APPROVAL

Name: Aaron Richard Nicholas Smith
Degree: Master of Arts
Title of Thesis: Textual Analytics and Personal Knowledge Management for Policy Analysis

Examination Committee:

Dr. Douglas McArthur, Chair

Dr. Nancy Olewiler, Senior Supervisor
Professor, Economics

Dr. Roman Onufrijchuk, Supervisor
Lecturer, Centre for Canadian Studies

Dr. Maria Teresa Taboada, Supervisor
Assistant Professor, Linguistics

Dr. Timothy R. Huerta, External Examiner
Assistant Professor, Rawls College of Business Administration
Texas Technical University, Lubbock

Date Approved: February 15, 2007
DECLARATION OF
PARTIAL COPYRIGHT LICENCE

The author, whose copyright is declared on the title page of this work, has granted to Simon Fraser University the right to lend this thesis, project or extended essay to users of the Simon Fraser University Library, and to make partial or single copies only for such users or in response to a request from the library of any other university, or other educational institution, on its own behalf or for one of its users.

The author has further granted permission to Simon Fraser University to keep or make a digital copy for use in its circulating collection (currently available to the public at the “Institutional Repository” link of the SFU Library website <www.lib.sfu.ca> at: <http://ir.lib.sfu.ca/handle/1892/112>) and, without changing the content, to translate the thesis/project or extended essays, if technically possible, to any medium or format for the purpose of preservation of the digital work.

The author has further agreed that permission for multiple copying of this work for scholarly purposes may be granted by either the author or the Dean of Graduate Studies.

It is understood that copying or publication of this work for financial gain shall not be allowed without the author's written permission.

Permission for public performance, or limited permission for private scholarly use, of any multimedia materials forming part of this work, may have been granted by the author. This information may be found on the separately catalogued multimedia material and in the signed Partial Copyright License.

The original Partial Copyright License attesting to these terms, and signed by this author, may be found in the original bound copy of this work, retained in the Simon Fraser University Archive.

Simon Fraser University Library
Burnaby, BC, Canada
Though many software tools for managing digital information exist today, their suitability for knowledge work in the domain of policy analysis is not clear. This thesis identifies and applies criteria for evaluating the suitability of knowledge management tools for policy analysis. Top ranked textual digitization and analysis tools from various disciplines are evaluated by comparing parallel analyses on a large, idiosyncratic collection of multi-disciplinary texts primarily drawn from Public Policy, Knowledge Management, and Cognitive Linguistics. Strategies developed for managing a large collection are discussed. Where current tools fall short, the thesis proposes and evaluates a multidimensional schema for automating analysis of large collections of multi-disciplinary texts. The schema’s consensus-reality basis is described, its philosophical premises are articulated, and its relevance to the fields of Computing Science, Linguistics, and Policy Sciences is explained.
DEDICATION

I owe a debt of gratitude to many thinkers and writers who have challenged and inspired me over the course of my academic career. In particular, I want to thank Benjamin Schneiderman of MIT and Stuart Card of Palo Alto Research Centre, whose fascinating explorations have inspired my own. Special thanks to Ryan Sadler and Nicholas Cherney, interlocutors in many long late night conversations that lead to and refined this work.

This thesis represents many years of thought, trial, and development. I dedicate this work to my loving wife Meagan. You believed in me even when the horizon obscured the finish line. Your support helped me to cut a swath directly towards that goal.
ACKNOWLEDGEMENTS

I want to thank Dr. Nancy Olewiler for her consistent support, and the Dean of Graduate Studies, Dr. Jonathan Driver, for affording me the opportunity to pursue my studies by Special Arrangements. To the latecomers to this project, Dr. Roman Onufruchuck and Dr. Maite Taboada, I extend my heartfelt thanks. Your generous contributions of domain expertise has improved my thesis immeasurably. And I particularly want to thank my external examiner, Dr. Timothy Huerta for his insightful and incisive critique—you trod bravely where many feared to! Thank you all for your help along the way.
# TABLE OF CONTENTS

Approval .............................................................................................................................................. ii  
Abstract .............................................................................................................................................. iii  
Dedication .............................................................................................................................................. iv  
Acknowledgements ............................................................................................................................... v  
Table of Contents ............................................................................................................................... vi  
List of Figures ......................................................................................................................................... viii  
Executive Summary ............................................................................................................................. i  

## Chapter 1: Introduction and Motivation ......................................................................................... 4  
1.1 Disciplinary Heritage ..................................................................................................................... 4  
1.2 Original Inspiration ......................................................................................................................... 9  
1.3 Ideal Formulation ............................................................................................................................ 16  
1.4 Intellectual Motivation .................................................................................................................... 22  
1.5 Traditional Disappointment ............................................................................................................ 30  

## Chapter 2: What are the Available Tools Good For? ..................................................................... 43  
2.1 Operational Assumptions ................................................................................................................ 43  
2.2 Demanding Environments ............................................................................................................... 44  
2.3 Evaluation Criteria .......................................................................................................................... 47  
2.4 Technological Tools ......................................................................................................................... 49  
2.4.1 Software Classification ............................................................................................................... 49  
2.4.2 Hardware Description ............................................................................................................... 51  
2.5 Suitability Evaluation ....................................................................................................................... 53  
2.5.1 Optical Character Recognition (OCR) ...................................................................................... 54  
2.5.2 Voice Recognition ...................................................................................................................... 55  
2.5.3 Bibliographic Reference Management ...................................................................................... 56  
2.5.4 Qualitative Data Analysis (QDA) .............................................................................................. 57  
2.5.5 Document and Meta-data Management ................................................................................... 59  
2.5.6 Content Management ................................................................................................................. 61  
2.5.7 Content Analysis, Semantic Network Analysis ......................................................................... 63  
2.5.8 Content Summarization .............................................................................................................. 65  
2.5.9 Natural Language Processing .................................................................................................... 65  
2.5.10 Software Virtualization ............................................................................................................ 66  
2.6 Tailored Solutions ............................................................................................................................ 68  
2.6.1 Best Fit ......................................................................................................................................... 68  
2.6.2 Practical Recommendations ........................................................................................................ 72  
2.6.3 Future Exemplar of Technologically-enabled Policy Routine .............................................. 74
LIST OF FIGURES

Figure 1. Levels of Acceptable Abstraction .................................................................36
Figure 2. Cynefin Sense making model ........................................................................37
Figure 3. Cynefin Decision making model ..................................................................38
Figure 4. Cynefin Ideal Knowledge Flows .................................................................39
Figure 5. Criteria Matrix for Software Suitability Evaluation ......................................47
Figure 6. Flow Diagram of Knowledge Management Tool Usage ...............................73
Figure 7. Phenomena of Human Experience ..............................................................102
Figure 8. Elements of Context situate Categories of Things .......................................105
Figure 9. Dimensions are Elements of Context that situate Entities ..........................109
Figure 10. Mental Extents underlie all others .............................................................110
Figure 11. Prototypical Instances of Dimensions and their underlying Extents ..........111
Figure 12. Dimensions of Existence .........................................................................112
Figure 13. Modalities are Elements of Context that prototypically situate States ........115
Figure 14. Properties derived from the definitions of various Modalities of Existence .................................................................................................................................117
Figure 15. Extents of Dimensions ................................................................................119
Figure 16. Related Instances of Orders by Dimension ...............................................121
Figure 17. Dimensional Requisites ............................................................................125
Figure 18. Binary-paired Properties serve to structure Child Instances of Categories .................................................................................................................................129
Figure 19. Customize My Documents .........................................................................144
Figure 20. Experiment with Folder Views to show only the Folders you have recently used .........................................................................................................................145
Figure 21. Make sure the Citation Information is Accurate .........................................147
Figure 22. Drag the Article from the Desktop to the my documents\eArticles\in RefMan .................................................................................................................................148
Figure 23. Follow the shortcuts and instructions. When complete, go to next step ......149
EXECUTIVE SUMMARY

This executive summary provides a brief overview of the thesis’ objectives. It clarifies the rationale for inclusion of each section and how they build upon each other in order to reinforce the connections between them.

My first objective is to explain how I have come to bring such apparently diverse and divergent disciplines together. This explanation manifests in Chapter 1 as a narrative account of my own encounter with various ideas and insights from the disciplines of Public Policy, Knowledge Management, and Cognitive Linguistics. My experience with these disciplines has provided me with the perspective to see how technological capacities developed in one discipline could be feasibly be applied to address practical problems in another, and how certain disciplinary theoretical advances might benefit practical solutions across domains.

Accordingly, my second objective is to see what existing textual analysis tools and knowledge management strategies can accomplish; and to identify which of them may prove useful in the daily work of policy analysts. Further to this objective, I will describe a concise framework that analysts may use to evaluate the usefulness of other textual analysis and knowledge management software. Where existing tools and strategies are well suited to the rigors of knowledge work in the domain of public policy, I offer qualified recommendations as to which tools to use. Where they are not, I catalogue their shortcomings as targeted grounds for future improvements to the software packages.
Given the diversity of information sources that policy analysts and other knowledge workers must daily evaluate and mobilize (Weimer & Vining, 1992a; Bardach, 2000b), my third objective makes this exercise practical, by sharing some ‘smart practices’ (Bardach, 2000a) I have developed for managing a large, transdisciplinary collection of texts, using the above recommended tools and practices. These practices are found in Appendix D, *Smart Practices for Managing a Large Transdisciplinary Collection*.

Since the most significant barrier to software recommendation is the time-intensive commitment required to successfully analyze texts using any given software package, measures taken to reduce analysis time costs — for example, by automating certain stages in the analytical process — would improve a the software’s suitability for policy analysis. Accordingly, my fourth objective is to argue that for large transdisciplinary textual collections, the best and most feasible improvement to existing textual analytic tools for knowledge management in policy analysis would be to automate file and meta-data management and content classification.

Automated classification of textual content usually requires a classificational schema upfront. Though artificial intelligence can be used to ‘learn’ and ‘evolve’ a classification system (Matuszek et al., 2005) most schemas are built to explicate a single given domain. This poses a problem for transdisciplinary collections. When automatically processing texts from different domains or disciplines, it is important to try to avoid building disciplinary bias into a categorization scheme. Ideally, what is needed is a classificational schema that describes ‘consensus-based reality’ (i.e., the generic and
generally agreed-upon version of the world that conceivably underpins any disciplinary ‘versions’ of reality) (Panton et al., 2006; Thompson, 2001).

In their originary domain of philosophy, such classification schemes are called ontologies (Fikes & Farquhar, 1999; Grecko, 1995; Guarino & Welty, 2004). Though each discipline may have different names and uses for them, (i.e., policy sciences call them Goeller scorecards or criteria matrices; cognitive sciences call them mental frames or image schemas; life sciences call them Linnaean classifications or biological taxonomies; library sciences call them card catalogues or subject headings; information sciences call them controlled vocabularies or website indices) ontologies are at essence a classificational system that can be used to describe things that exists in a given domain using a comprehensive yet parsimonious set of classes or categories.

Since the literature on cognitive linguistics affirms that the ambiguity of linguistic word sense generally requires a reader to resolve an appropriate meaning (Empson, 1998; Croft & Cruse, 2004; Croft et al., 2004; Day & Lord, 1992; Rockwell, 2005b), and given that readers actively construct linguistic meaning with reference to various levels of textual and real-world context (Fauconnier, 1998; Fauconnier, 2000; Werth, 1999c), attempts to automate textual analysis and classification require methods to reduce word-sense ambiguity and capture the necessary context (Curtis, Cabral, & Baxter, 2006).

As a step toward improving existing tools by facilitating automated content classification, my final objective is to present a schema that describes both the things that exist, and the context within which they exist. Ideally, once implemented, these contextual categories could offer a way forward toward the multi-layered thematic analysis of qualitative data described by Guest & Mclellan (2003).
CHAPTER 1: INTRODUCTION AND MOTIVATION

1.1 Disciplinary Heritage

This thesis is the culmination of three years spent exploring several related disciplines. I began my graduate career studying Public Policy with an Honours degree in Communication. I have since expanded my course of study by Special Arrangements to include Knowledge Management with a minor focus on Cognitive Linguistics. Upon completion, I will have carved my Masters degree from the margins of several interdisciplinary academic traditions.

I use the term ‘margins’ to situate my research vis-à-vis my understanding of the primary objectives of the various disciplines implicated in this work. Though my project is housed in the Applied Sciences, I have focused my studies in various complimentary domains of Social Science, with only a few brief forays into the Natural Sciences.¹ As these domain distinctions exist to segment the modern University, I will assume that all share something of a commitment to teaching.

For example, I would argue that the broad aim of Applied Sciences is to train practitioners in the systematic application of scientific and human knowledge in various

¹ I want to understand the rudiments of neurological and electrochemical bases of brain function; not to be the final word on cognition (or the mental, emotional, and motivational process more broadly) but rather, to better inform my development of a human-centered heuristic for ‘how we think’ and ‘how we come to understand’. To be clear; this is not a treatise on biochemical brain function, or a detailed account of the neural pathways in the neocortex; this is the development of a user-centered model primarily informed by personal accounts of the process and methods used to approach and manage work in an information-saturated environment. Since my readings have taken me into so many domains in which I am admittedly no expert, my attempts to characterize my understanding of their various ‘primary objectives’ must be considered a lay perspective.
problem domains. Were I required to surmise a unique pedagogical purpose for each of
the applied domains I have drawn upon, I would have to hazard a guess for the Policy,
Management, Library, Information, and Computing Sciences. In a similar vein, I would
argue that the didactic goal of the Social Sciences is to train researchers in the systematic
theorization, discovery and analysis of human knowledge across social domains. Those
most central to this work include Organizational and Communication theory, Philosophy,
Linguistics, and generically also Social Science Research Methodology. Though furthest
away from the heart of my research, I would surmise that the educational purpose of the
Natural Sciences is to train scientists to theorize, test, and discover scientific knowledge
across physical domains using scientific methods.

Now, while a certified expert from any of these fields may find my ideas about
their primary objectives to be bland, overly generalized, or even vacuous, I insisted upon
including them for two reasons. First, they serve to highlight what this thesis is not. It is
not an attempt to suggest my expertise in each of these fields. I cannot possibly be aware
of each detail of every perspective within all these disciplines. (Referring to notes 2-4

Accordingly, I imagine that the goal of Policy Sciences is to train analysts how to marshal a variety of
methods to characterize and operationalize problems, identify and evaluate options, and to qualify and
recommend solutions. The object of Management Science (Business Administration) is to train managers
to strategically direct the operations of organizations towards greater effectiveness and increased
efficiencies. The purpose of Library and Information Sciences is to train librarians and information
architects to create and implement logical systems to index the content of published materials and manage
access to valued repositories. The object of Computing Sciences is to train engineers qualified to design
and build hardware and software technologies.

In my view, the goals of Organizational and Communication Theory (and Research) appear to be more or
less parallel; they aim to advance understanding of the dynamics that govern (respectively) organizational
and communicational success across various contexts; and to train management consultants or public
relations experts. As for the other core contributors to my work; Philosophy aims to advance theoretical
thought on the nature of reality and to train critical thinkers, while Linguistics sets out to describe the
universal features of language in a system of rules, or grammar, and to train language experts. Social
Science Research Methods theorizes, implements and refines systematic methods for discovering,
analysing, and publishing social data from sources that may not be suited to the classic scientific method.

Cognitive Sciences seek to advance understanding of cognitive functions broadly across species.
Neurobiology delves into the physiological and biological substrate of the brain, while Psychology focuses
on the level of cognitive mechanisms and processes of the mind.
below, in some cases it should be obvious that I have only the vaguest ‘folk-theoretic’
notions of what goes on in a field.)

Of course my readings alone do not qualify me as a policymaker, corporate
executive, librarian, software engineer, neurobiologist, psychologist, or cognitive
scientist. Nor are they sufficient to classify my work as belonging to these domains. The
applied nature of my research could place me closer to qualifying as a policy analyst,
communication scholar, organizational theorist, philosopher, and possibly even as an
honorary discourse linguist. But while my research draws more deeply upon these areas,
it neither conforms to their idealized academic forms, nor attains to their ostensible
disciplinary goals.

For example, while my thesis centers on Knowledge Management, I have
abandoned its traditional unit of study, the organization, in favour of a course of study
focused at the level of the individual. 5 And while Policy Sciences is properly about
training the analyst in the methods and process of doing policy analysis, I am more
concerned with understanding the operational dynamics and constraints within which the
analyst works, and on finding tools that can help them excel in these conditions. 6

I wanted to highlight these facts from the outset. An expert’s opinion will almost
certainly find significant room for improvement in the form or content of the exposition
that follows. Wherever possible, I have integrated the feedback and comments I have
received from my colleagues, mentors, and advisors. While this support covered a wide
range of expertise, some areas remain solely a synthesis of my studies and thinking to

5 This choice was made on both practical and philosophical grounds. These rationales are discussed in
Sections 1.2 and 1.3, and further in Chapter 3.
6 The rationale for this focus follows on pp 4-6.
date. Of course, any limitations in this work are my own. Of staunch disciplinarians I ask
only that the critical instinct be moderated by an appreciation of the ultimate purpose of
this work, for though I welcome all critique and relish the opportunity to refine and adjust
my understanding of things, I have also accepted the limitations of scope imposed by the
genre, and cannot give all problems the consideration they certainly deserve.

My second reason for including my generic characterisations of disciplinary
objectives is just that they are a fairly reasonable example of an attempt to typify or
generalize the properties of loosely related categories with language. Classification is a
neat, reductive way to address complexity simply. At its core, this is just what this thesis
is about (see Chapter 3). Each discipline is in fact a massively complex domain of
knowledge, with a history rich in conflict, rife with turf-wars, and prone to much name­
calling and hand-waving. It is difficult to succinctly describe this kind of complexity
without getting into too much detail about their various traditions, approaches, and
schools of thought. Though all disciplines retain their own unique topical focus they also
share certain characteristics. These are the ostensible bases of the foregoing
categorizations.

Pragmatically, this thesis recounts my attempts to grapple with the problem of
acquiring, understanding, classifying, and mobilizing the content of discourses that I
am not necessarily familiar with. This is a significant problem facing individual
knowledge workers and interdisciplinary teams today: I am not an expert in this area, but
I need to know this for my work. Put another way, this thesis has been a process of
systematically observing and attempting to technologically operationalize\(^7\) my learning process in a context where I am missing the full disciplinary background usually considered necessary to developing high-functioning understanding of the issues.

The practical intention of my thesis is to offer the professional Policy Analyst – who is taken as a qualified exemplar of an interdisciplinary knowledge worker facing just such a problem – a range of tools uniquely suited to their professional operating conditions that will augment and expand their capacity to capture, categorize, analyse, and recontextualise the various streams of information required to excel in the daily execution of their tasks and responsibilities. My analytic task is to identify and evaluate technologies that show promise in this regard and to catalogue their short-comings in order to recommend improvements targeted for the Policy Analyst. My contribution to theory centres on the development and description of a sparse schema\(^8\) for building rich models of the complex realities of things and their context using automated analysis of digital texts from any discipline (Werth, 1999g). The classificational schema must therefore be experientially grounded, discipline-neutral, philosophically explicable, universally applicable and infinitely extensible. I hope that its future implementation and automation in discourse-analytic personal knowledge management software will begin to facilitate true transdisciplinary understanding among lay readers and deepen understanding and strengthen collaboration on interdisciplinary teams.

---

\(^7\) By technologically operationalize, I mean to identify and evaluate software tools that facilitate the various stages of learning, mobilizing and applying new (and existing) knowledge.

\(^8\) The term 'schema' can be understood variously as a rubric, coding system, ontology, classificational system, and controlled vocabulary. In each case, idea behind it refers to a structured representation of knowledge.
The four academic disciplines most central to these purposes – Policy Analysis, Knowledge Management, Communication, and Cognitive Linguistics – are by nature interdisciplinary fields of study. This strange phenomenon of multiplexed interdisciplinarity – an interdisciplinary course of study of interdisciplinary traditions – goes a long way toward explaining the breadth of my readings. What remains to be accounted for can be explained by my intent commitment to meta-observational learning while attempting to simulate the diversity of informational sources that is common in policy analysis.

While the exact formulation of my thesis has changed over the course of the last three years, at its core it has remained the same: If and when existing tools do not meet the needs of users, the bases of this failure should be examined in the light of theoretical and applied knowledge from relevant domains in order to refine and tailor tools to the needs of users. This document is, as much as anything academic, a narrative account of my progress in this regard. It is a record of what I have learned as I have attempted to better manage the knowledge I have gathered over the course of my research.

Let’s start from the beginning.

1.2 Original Inspiration

Reading Paul Werth’s monograph, Text Worlds, five years ago as an Honours undergraduate student in Communication changed the course of my graduate career. My

---

9 Policy Sciences draw from Political Science, Economics, and various Social Science Research Methodologies. Knowledge Management grew out of the intersection between the fields of Management, Information, and Computing Sciences. Cognitive Linguistics explores the connection between cognitive and language function in humans, drawing from Cognitive Sciences, Psychology, and Linguistics. Communication Studies, one of the original flagship interdisciplinary disciplines, draws from History, Political Science, Sociology, Critical Theory, Philosophy, Technology Studies, among many others. 10 Since policy analysts must constantly process information from sources in unfamiliar territory, my eclectic collection is an approximation of these conditions.
introduction to Cognitive Linguistics\textsuperscript{11} – a loose federation of the Cognitive Sciences, Psychology and Linguistics – proved both accessible and fascinating. Werth artfully provided the pedigree of every significant concept he invoked or developed, effortlessly folding interdisciplinary history and glossary into his easy expository flow. Though I was at the end of my undergraduate career, I was eager to learn more.

*Text Worlds* introduced me to some of the thorny problems surrounding knowledge representation and modelling of textual discourse. Prior to his death in 1997, Werth developed a frame-based schema for manually visualizing ‘text worlds’ – the descriptive ‘worlds’ that authors create for readers – by mapping lexicographical cues, like parts of speech and verb tense, into spatially and geometrically symbolic areas in a diagram. The idea is that our ability to understand things in text worlds mirrors our ability to understand our experiences of the real world.\textsuperscript{12} His schema is therefore a heuristic for both.

While his mapping schema is robust and elegant, its implementation is arduously manual and requires a degree of lexicographical expertise at the level of a PhD candidate, a professor of literature, or a skilled linguist. In its present form it is better suited to academic expertise and timeframes than for practical real-world application. Indeed,

\begin{itemize}
\item \textsuperscript{11} Whereas mainstream structuralist or generative traditions study linguistics more or less as an intrinsic system of rules, cognitive linguists study language as an artefact of mental activity and approach its grammar, pragmatics and semantics with reference to models of human cognition.
\item \textsuperscript{12} Werth argues that the cognitive capacities that enable us to understand authorial description and action are patterned after the cognitive rules that govern our ability to understand our circumstances in the real world.
\end{itemize}
today the application of text-world theory is almost exclusively the province of literary scholars.¹³

But more importantly, reading Werth’s arguments convinced me that a generative¹⁴ approach to language cannot provide a compelling account of linguistic meaning. Though Werth’s earlier work as a discourse grammarian was rooted in that tradition, the implication of his theories on coherence lead him to realize that Generative Linguistics “excluded almost everything that did not fall into a very restricted area of syntax and phonology,” and that excluded discourse (1999a).

While developing a mechanism to explain how coherence actually works (1984), Werth began to look at the phenomenon beyond the generative focus on sentence grammar. It became apparent to him that “verbal context alone was not enough either to characterise discourse or ultimately to explain how discourse elements interact with each other” (1999b). In Werth’s view, the missing ingredient was a fuller notion of context.

While generative linguists ignore it on principle, insisting that it is impossible to model mathematically, Werth insists that however difficult, factors that are central to so many issues in human language cannot be so easily dismissed.¹⁵

The sceptical view is correct to insist that context must have limits. However, to follow in the generative line of reasoning with a straw man definition of context as

¹³ Nevertheless, the possibility of one day augmenting and implementing Werth’s schema as a software tool for visualizing textual and discourse level semantics has long since captured my imagination.

¹⁴ The generative theories developed by Chomsky are prototypical of the tradition.

¹⁵ Consider this extended quotation by Werth (1999b)
I firmly believe that we cannot afford to ignore the context, not least because it is the source of very many of the problems thrown up by current approaches to the study of language. These include problems in pragmatics such as relevance and speech acts – indeed, we might say that the whole of pragmatics is located here – problems in semantics, such as reference, classification and inference, as well as problems like anaphora and definiteness, often treated as purely syntactic, and problems to do with stress and intonation, often treated as phonological or even phonetic matters.
'everything about everything' is recipe for disaster. The real debate begins with the question of how to define and delimit context. Werth initially describes the context of a piece of language broadly as its 'surrounding environment.' More specifically, he defines verbal context as the "language surrounding a particular sentence or proposition" and situational context as consisting of "the immediate situation and cultural background surrounding the language event under scrutiny." (1999e)

Though a rational observer might agree that these seem to be sensible definitions, in both cases it remains to be decided how to restrict the notion of context in some principled way. The principle Werth invokes is premised on his understanding of the relationship between discourse and text. He defines discourse as an actual, complete stretch of language that occurs in a real setting with a perceived beginning and end and that is comprised of a series of related utterances, or single situated speech acts. The function of discourse is to encode the particular situation it refers to in the form of a text (1999e). So a discourse is a text in a physical setting, much as an utterance is a sentence in a real context. A text is a discourse-encoded representation of situational context. The discourse and utterance are the real thing, whereas the text and its sentences are abstractions away from the real thing.

This abstraction does not trouble the discourse linguist in the least. According to Werth, the discourse producers include each element of the immediate situational context that they feel is necessary to achieve their communicational purposes. The text is therefore a perfectly succinct delimitation of the relevant situational context. In Werth's own words:
The relevant situation is ... precisely restricted in every case by the text which its discourse produces, and the discourse provides just enough detail to set up a text world and to activate the relevant areas of knowledge. No further details are necessary than are provided by the discourse together with information available from the frames accessed by participant knowledge. (emphasis added) (1999a)

Werth’s’ theory stipulates that while texts contain all the details necessary to understand them, these cues require the reader to know certain things that the author considers ‘common knowledge.’ These textual details serve to activate the readers memory, including “complex conceptual structures known as frames” (1999b). In a linguistic sense, a frame is a prototypical description which sets the working conditions within which a term is used. (1999e) From a cognitive perspective, “frames are whole chunks of experience and situations, codified and stored in memory as single items (see Werth, Chapter 4)” (1999b; Lowe, Baker, & Fillmore, 1997; Baker, Fillmore, & Lowe, 1998; Fillmore, 1997; 1998; 2000; 2001; 2002; 2002a; 2002b; 2003; Narayanan, Fillmore, Baker, & Petruck, 2000; Narayanan, Petruck, Baker, & Fillmore, 2003; Petruck, Fillmore, Baker, Ellsworth, & Ruppenhofer, 2004). In either case, frames are continuously adapted and added to as new experiences and details are associated with existing knowledge. (1999f) Werth contends that frames function to ‘flesh out’ the discourse from the knowledge and imagination of the participants. (1999b)

The reading process Werth postulates follows a conversational or discursive pattern that ultimately implicates a dynamic range of cognitive capacities and activities. As a process of negotiation, it depends on the existing knowledge of the reader, their ability to follow the text’s symbolic references and effectively associate them with their own prior knowledge. To the non-linguist, this might seem fairly basic and common-
sensical. However, for the generative linguist,\textsuperscript{16} this strays into the uneasy territory of
‘semantics’—study of the relation between the signs and the objects to which they apply—and into the downright disreputable ‘pragmatics’—the study of the relationship between sign systems and the user. (Croft, 1991; 2006)

Yet Werth remained unabashedly unapologetic for this association. His thesis situates the entirety of these two discrete areas of linguistic inquiry into two adjacent, layered mental spaces or ‘worlds.’ Werth argues that all language happens in the context of a real\textsuperscript{17} situation that is represented (in the mind) as the ‘discourse world.’ Our perceptions and our knowledge about the objects we perceive constitute the basis for the discourse world. Inasmuch as interpretation colours our perception, the discourse world can be considered a construct, though ostensibly one based in reality. He argues further that use of language also presupposes “the existence of a conceptual domain of understanding, jointly constructed by the producer and recipient(s) (1999b).” Though this ‘text world’ is also backed by participant knowledge, it is a ‘total construct’ that is

\textsuperscript{16} Generative linguists tend to focus on syntax—the study of the relations between signs—in an attempt to describe a universal grammar for language.

\textsuperscript{17} Despite a distinctly phenomenological attitude towards human knowledge of reality, his ontological ambivalence does not hamstring his theory of language. No matter how slippery, we ‘readily accept’ the assumption that reality exists. However, in Werth’s view, the connections between the sign and ‘reality’ are not built in to the sign, but rather stipulated by the discourse.
entirely dependent on resources of memory and imagination, rather than on direct perception.\textsuperscript{18}

My reading of Werth introduced me to a number of seminal ideas which have redirected the course of both my intellectual development and my academic career. These, in particular, have served to frame my approach to my graduate studies:

1. Language is a structured system that requires active participation to work.

2. Meaning is not innately embedded in Language; it is created by readers.

3. Meaning depends on the activation of necessary knowledge in memory.

4. The cognitive processes of associative learning and long-term memory allow for subtle variations of interpretation, while shared culture, common socialization, and similar neuronal function serve to normalize interpretive licence and ensure some degree of effective communication.

5. The relationship between sign systems and the user determines the relation between the signs and the objects to which they apply with reference to the relations between signs. (Pragmatics determines semantics with reference to syntax.)

\textsuperscript{18} Consider Werth's thesis in his own words: "My main thesis is that all of semantics and pragmatics operates within a set of stacked cognitive spaces, termed 'mental worlds'. Connections with 'reality' are stipulated rather than built in, and indeed the very notion of reality itself is an assumption, which we readily accept, but have no direct access to. My argument for this position, in a nutshell, is that uses of language presuppose occurrence in a context of situation, and that on top of this they also presuppose the existence of a conceptual domain of understanding, jointly constructed by the producer and recipient(s). The former is the immediate situation (the representation of which I will henceforth call the discourse world) and the latter a text world. The discourse world is based on perception, backed up by knowledge of the elements perceived. To the extent that it is founded on interpretation, and is 'filled in' and edited, the discourse world is a construct, although one which we may suppose is founded on 'real' external circumstances. The text world, though, is a total construct, so therefore negotiated by the participants through the medium of the discourse, again backed up by relevant knowledge. Since it is a construct, it is dependent on resources of memory and imagination, rather than direct perception. (Werth, 1999b)"
1.3 Ideal Formulation

As these ideas have percolated over the years, they have developed into goals. As I read a text, I want to be able to see the various levels of structure that I know exist in it. To start, this includes the part of speech for each word (N, V, ADV etc.) and phrase (Noun Phrase, Verb Phrase, Prepositional Phrase etc.), as well as word-level morphological (tense, number, etc.) and syntactic features (Subject, Object, Direct Object etc.). But texts are embedded with many more levels of structure, from the logical structure of its propositions, to its argumentative, rhetorical, narrative, and discourse structures, and finally to its document-level layout and formatting structure (MacRae, 1993; Exegenix Inc, 2004).

These structural features are the closest thing the reader has to the objective facts of a text. The appearance (or non-appearance) of words in a text is a fairly objective fact, as is their part of speech, tense, agreement, aspect, etc. along with other details of their morphological, syntactic or grammatical structure. Experts may differ as to which category applies or how best to make that determination, but generally agree that there is a sense in which belonging in these categories is, well, categorical. It can be assessed as right or wrong; correct or incorrect. Most of these linguistic facts are difficult, but not impossible to determine reliably using existing, available computational methods (Adriaans, 2001; Fletcher, 1999; Heemskerk, 1993; Kintsch, 1999).

The upshot is that if texts contain objective facts, then we can expect that these facts ground the otherwise subjective processes of reading, interpretation, and the formulation of meaning. The words used in a text mean something specific, and though the intended sense of a term may be uncertain to any given reader, the range of possible
interpretations is not infinite. Grammar constrains and “guides us along rich mental paths, by prompting us to perform complex cognitive operations” in an effort to understand textual meaning (Jannsen & Redeker, 2000; Werth, 1999c).

But texts are a lot more than instances of layered symbolic structure. They contain many types of relations and evoke many more. For example, each word in a text has a set of lexicographical relations with its definition, which often include a range of possible senses for any given term.19 In linguistics, this phenomenon is called polysemy. Figuring out which sense is most appropriate, while difficult to automate, is something knowledgeable language users do almost ‘without thinking’ (Jannsen et al., 2000). In Werth’s theory, readers use contextual cues that activate their memory of prior knowledge to do what computational linguists call ‘word sense disambiguation.’

Where terms are not explicitly defined in a text, lexical relations are ‘evoked.’ On the other hand, the relation a word has with the words immediately around it (its verbal context) is an ‘embedded’ proximal (or sequential) relation. But words don’t have to be beside each other to be related. Language also provides ways to refer to ideas, actions, and things that were mentioned previously in a text. Linguistics calls this phenomena ‘anaphora.’ Intersentential anaphoric reference requires the reader to resolve the reference, so is considered to be evoked (though the object of the reference is embedded in the text). Other evoked relations include various semantic relations of terms with their parent concepts, child concepts, equivalent concepts, opposite concepts, etc. While not necessarily embedded in the text, these and other relations form a semantic network that

19 The number of possible senses for even simple sentences can be surprisingly high (CITE).
ostensibly allow for user to make inferences when faced with new terms or novel linguistic constructions (Halliday & Hasan, 1976; Danzico, 2006).

Gilles Fauconnier, one of the founders of the Cognitive Linguistics research program, reaffirms Werth's contention that "any language form in context has the potential to trigger massive cognitive constructions, including analogical mappings, mental space connections, reference point organization, blends, and simulation of complex scenes. (2000)." He points out how the "extreme brevity of the linguistic form" stands in sharp contrast to the "spectacular wealth of corresponding meaning construction (2000)." Though every reader responds somewhat differently to the grammatical structure and embedded cues of a text, the miracle, he points out, is that effective communication is nevertheless still possible (2000).

Fauconnier argues that the methods employed by cognitive linguists "must extend to contextual aspects of language use and to non-linguistic cognition. This means studying full discourse, language in context, inferences actually drawn by participants in an exchange, applicable frames, implicit assumptions and construal, to name just a few (2000)." Even prior to reading his call to arms, I had already begun to imagine what Werth's schema and methods might look as a software tool. Of course, no such software exists. However, as I considered the requirements for such a tool, I began to find that certain core elements and functionality do exist in other software today.

On a practical level, Werth focused my attention on finding or creating a system capable of capturing both verbal and situational context, that could eventually allow me to model, simulate, observe, and test theories regarding the activation of the 'frames' of common knowledge required to understand any given text. As I understand them,
Werth’s theories would require capturing not only the text in its entirety, but also various salient details about it, like its bibliographic and document-specific meta-data, its formatting and layout, as well as its cultural role and relative location vis-à-vis other texts in a larger discourse.

Such a tool would require a range of textual analytic capacities, drawn from natural language processing (NLP), computational linguistics, and computing science more broadly, as a necessary precondition to the kind of discourse processing described by Werth. The software would have to treat the text to a range of analytical routines, in effect ‘pre-process’ it, before the reader approached it. These steps would include:

1. Parsing and Chunking: It would parse the text to identify sections, paragraphs, sentences, phrases, words and punctuation.

2. POS-tagging, Morphological Analysis, and Word Sense Disambiguation: It would identify parts of speech and attempt, initially by discarding non-grammatical possibilities, to disambiguate phrasings that could be interpreted several ways.

3. Content Analysis, Word Stemming, and Lemmatization: It would have the capacity to identify all the words and phrases the text contained, to aggregate and enumerate concepts by stemming or lemmatization, (identifying the morphological root of a word; like concept* for conceptual) and to find both particularly common and especially unique words, phrases, and concepts.

4. Document Structure Tagging, Semantic Mark-up: It would intelligently and automatically assess the structure, format, layout, and embedded objects of a text to identify and extract its Chapter Titles, Sections, Authors, Dates, Page Numbers,
Images, Diagrams, Equations, Charts, and Tables and then automatically normalize their formatting.

5. Automated Entity Recognition: It would be programmable to recognize and extract a variety of ‘entities’ from a text, like proper names, company names, addresses, phone numbers, but also DOI, ISBN, bar codes, academic references, citations, foreign language phrases (Teller, 2000; Gildea & Jurafsky, 2002; Jurafsky & Martin, 2000).

With this kind of pre-processing in place a researcher could generate a ‘cold’ dictionary-defined text world that contained an automatically generated, basic collection of related frames required to ‘understand’ the text. It would automatically map each word into a frame, using lexicographical and morphological cues, like parts of speech and verb tense, to dynamically situate and populate the frames with reference to each other, as described by Werth’s schema.²⁰ (Frank & Semecky, 2004)

Researchers could then observe as readers from different backgrounds adjust and refine the proposed frame contents and arrangement and begin, over time, to populate them with their own associations and inferences. The theory goes, at each concept in the text, the reader decides if this concept is one she already knows about, one that she has already heard mentioned so far in this text, one that is somehow related to either of them,

²⁰ In Werth’s schema, every word refers to a concept and has a frame. Any of the semantic relations possible between concepts can be contained within a given frame. An activated frame contains every encountered relation of a given concept along with any of its related, activated concepts. So though every word has its own frame, some frames contain more relations and frames than others, due to variations in their activation level.
or maybe even one she doesn’t know about yet. The reading process of real readers could be compared to the simulated neural-network ‘model reader’ to refine the model.\textsuperscript{21}

The idea is that the software identifies as many of the different kinds of facts about the text as is possible and assigns them to layers that usually remain hidden to the reader, but which facilitate both further discourse analysis and the reader’s ability to interact with the text in several new ways. Primarily, this includes dynamically visualizing the narrative frames of knowledge explicitly invoked by the text, together with the means to create and interact with those frames implicitly evoked in the reader.

To the extent that a reader’s existing knowledge is implicated in the process of formulating textual meaning, texts, the concepts they contain, and the relations they evoke become a proxy for studying existing knowledge. In this view, knowledge is not a thing, but a process of relating what we recall of what we have encountered previously with what faces us now (Snowden, 2002).

The literature on knowledge has developed and discarded many metaphors to describe knowledge. Armbrecht (2001) describes it as a flow; Richmond (1988) describes it as a force field; and Landry (2001; 2005) describes it as exchange. However, these metaphors are primed for discussing knowledge in an organizational context. This thesis follows the constructivist tradition of Glasser outlined by Riegler (2005). While no individual can become knowledgeable alone; the process of actively formulating, rather

\textsuperscript{21} Rather than a simple, word-based proximity model of activation used by older neural-networks, this model would scale to include phenomena like deixis and anaphora that stalled or were simply ignored by previous analyses. Previously, one neuron would fire for each word within a window of N-words, as the window passed over the text at some constant rate (Inc, 1993c; Inc, 1993b; Inc, 1993a; Inc, 1998b; Inc, 1998a; Woelfel & Richards, 1989; Woelfel & Stoyanoff, 1993; Woelfel, 1998a; Woelfel, 1998b; Woelfel & Barnett, 1992). The simultaneous firing of words together was due solely to their proximity and linear sequence. In this model, neurons would fire in a sequence following the arrangement of the text-world frames, so that whenever entity or actional anaphora appends new details to a previously activated frame, its neurons would refire.
than passively receiving the contents of individual knowledge is what primarily concerns this discussion (See Section 3.3 Epistemological Implications.)

If readers really are essential to the creation of the meaning of the text, then automated attempts at word-sense disambiguation and text-world frame activation will need correction. But if readers really are essential to the creation of meaning, they should not be bogged down trying to manually map out their process of comprehension. Accordingly, the software should automate as much of the process as possible and allow readers to easily audit their knowledge/frame-base as required.

To reiterate; this software does not exist, though I have found that (more or less) each of its required components do. Since I could not theorize and test the activation of knowledge in readers with a tidy software package, I began to try to lay down the groundwork that might allow me to do so one day. Practically, this meant starting to gather the texts that I encountered and beginning to organize and analyze them with the tools I could find. This amounted to **systematizing a collection process** (See Appendix D, Smart Practices), and **evaluating a range of software packages**. My progress, and a typification of the tools available for these purposes, is carried forward in Chapter 2.

### 1.4 Intellectual Motivation

My primary motivation for this thesis draws upon a unique set of philosophical perspectives on practical matters. By education or disposition (or possibly both) I have developed a pragmatic idealism about the prospects for positive change in the real world of human choice and action that is balanced by a healthy scepticism about the capacity of human organizations and institutions to adapt progressively to non-threatening priorities.
I have always imagined that an empowered approach would strive to understand the big picture by learning how good decisions are made by individuals and progressive directions are taken by organizations. Three years ago I set out on this path when I began my graduate career studying public policy. Policy analysis is a professional activity that employs a variety of methods to characterize and operationalize problems, identify and evaluate options, and to qualify and recommend solutions. From there, it is up to decision-makers to decide what to do and how to get it done.

Now, despite some evidence to the contrary, I do not innately trust decision makers to make high quality (reflecting the diversity in the ethical, moral, and pragmatic perspectives of their constituents) decisions any more than I trust human institutions to spontaneously adapt progressive (socially just, environmentally responsible) priorities absent a significant threat. Nor do I think that a reductive rationalist model of human decision making and organizational process literally describes what goes on in either case (Smith, 2001; Kurtz & Snowden, 2003). As Wildavsky put it, “Policy is a certain reflection of the political process and therefore must be embedded in the political process; ergo, it will surely and unquestionably be affected by political events.” (Wildavsky, 1964) But by learning about how things are done in the halls of power, I hoped to find room for improvement in individual choice and opportunities for leverage in human institutions.

I believe that studying policy analysis will serve to accomplish both. True, the existing system segments the development of rationales from the selection and implementation of policy options and consequently builds in many junctures at which policy decisions can go wrong (Nagel, 2002a; Dunn, 1981c). But analysts are trained to
be aware of these pitfalls and to produce policy advice that accounts for these systemic features (Weimer & Vining, 1992b; Geva-May, 1997a).

Realistically, however, analysts cannot exhaustively project the implications of the variations introduced along the process of policy making. They operate subject to the law of unintended consequence (Gibson, 2003). Though policy analysts are not the most central figures in the decision-making process, they do play a critical catalyzing role in the formulation of viable solutions to real problems.

While the overarching methodology used in policy analysis for structuring problems, ranking options, and selecting solutions is not strictly scientific, it is certainly pragmatic. The problem domains explored in policy analysis are described variously as squishy (Strauch, 1975; Carley, 1981), fuzzy, noisy, messy (Patton & Sawicki, 1993), or wicked (Ritchey, 1998) in that they usually include a complex of individual, social, organizational, environmental, and larger political considerations (Dunn, 2001a). Its informational sources are legion. They include raw numerical data, complex statistical and economic calculations, graphs and charts, published books and articles, internal policies and memos, personal correspondence and conversations, official statements and speeches, original research and polls. (This list could go on.) The pragmatism ascribed to practitioners in this field is most obvious in their response to this noise and in their default manner of dealing with informational deficits and surfeits (See Section 2.2 Demanding Environments).

---

22 Dunn (1981) describes the range of methodologies available to the analyst at each step in this process.
It is this pragmatism that leads me to believe that, in principle, policy analysts would welcome the introduction of a sufficiently low cost means of improving their capacity to manage noisy situations and multiple streams of information.

To my mind, certain fundamental premises of my thesis just boil down to common sense. Show someone how their tools are maladapted to the task at hand, provide them with an alternative that works, then sit back and watch the age-old drama of technology adoption. To formalize the proposition somewhat, I argue that the motivation of a rational knowledge worker towards increasing their own personal analytical capacity can be changed from command and control-style enforced 'compliance' of previous generation knowledge management (KM) initiatives (Malhotra & Galtetta, 2005) or an ideal of purely altruistic 'organizational goodwill' to a the rationale of highly practical individualist maximizer (Gauthier, 1996; Stone, 2001; John, 1998b), if and when a given tool's immediate value proposition outweighs its cost of adoption (its acquisition, implementation, and ongoing maintenance). Simply prove a tools' worth and its adoption need not be forced or feigned.

There is some evidence to support this contention. Reporting on a recent study of state level health policy makers, Sorian and Baugh (2002) suggest that those most likely to use digital resources in their daily work are the youngest. Of respondents under the age of 30 (most likely to be legislative staff), 89 percent said that they either read electronic copy more frequently or read electronic and hard copy an equal amount. In contrast, 88 percent of those over age sixty (most likely to be legislators) read hard copy more often (2002).
On average, policymakers said that they read “27 percent of what they receive for detail, skim 53 percent for general content, and “never get to” 35 percent of the material” (2002). They estimated that 49 percent of the information they receive is not relevant to their current work (2002). When asked what additional information services or materials would be helpful to them in performing their jobs, “one in five said that they could use more “empirical data,” [though] most focused on the need for better tools to navigate the information system” (2002). This is, however moderated by the fact that “27 percent said that they either get too much information or have no additional information needs” (2002).

The responses were drawn from a small study (N=292) in which the number of ‘young’ policy analysts (under the age of 30) is very small. These facts have the effect of moderating my optimism somewhat. They refocus my estimation of the perceived usefulness of a tailored collection of knowledge management tools for policy analysts. While most established policymakers may find only limited use for such tools, younger analysts would likely find them more useful.

I assume that those who rely exclusively on hard-copy lack a quality possessed by those who don’t, who might be described as ‘technologically-savvy.’ Though as a group, younger analysts may be more tech-savvy, some analysts from every age group are sure to qualify. And savvyness alone cannot predict a willingness to invest in learning how to make a system work and commitment to integrating processes in the daily routine (which I describe together as motivation). So technologically savvy, motivated, and likely younger analysts might find these tools useful. I think it is reasonable to suggest that
generally speaking, students of policy analysis would fit the profile of young, tech-savvy and motivated.

These findings could suggest that a system that scales to include traditional hard copy could broaden its attractiveness across age groups. In any case, tools which helped to filter out non-relevant information (by automatically digesting, categorizing, organizing, and summarizing) would likely broaden appeal. According to some analysts, navigational tools for negotiating the information system and targeting the discovery of empirical data would also prove helpful in job performance (2002).

Now while these data provide something of a pragmatic basis for my work, I began this section referring to my philosophical motivations. Ultimately, I want to help make better decisions that result in better outcomes across the full spectrum of human values. Coming from an undergraduate degree in the Social Sciences, I want to believe that decisions made without understanding are poor and that ‘understanding’ shy of comprehensiveness is shallow. But I acknowledge that the pragmatic need to make decisions and take action will never wait for perfect or even near-perfect information.

We live in an information-rich yet meaning-poor world. The fact of abundant information does not mean that we have access to it, that we understand it, or that we know what to do with it. ‘Smart’ decisions are made on the basis of more comprehensive understanding, not just with more information (Bardach, 2000a). Now, while there is a long tradition of inquiry in the Social Sciences investigating the use (or disuse) of Social Science information by policymakers, it is not my intention to present a universal solution to this problematic dynamic. I do, however, suggest that analysts should start with what they know and they should strive to understand more.
I propose that analysts who are able to dynamically search and efficiently recall their own knowledge as a starting point to wider information-seeking will be more effective in their work. They will become the person to whom their superiors turn when they need someone ‘who knows, or knows where to find it.’ An analyst who can take what they know and turn it into a structured repository that grows over time will come out ahead of those who pile up photocopies of unread articles on their desks.

While the costs of exhaustive research (the collection and analysis of all relevant information) are too high and its timelines are too long, good policy analysis still depends on good information. But the criteria for evaluating the ‘quality’ and ‘relevance’ of information are likely to change over time and according to the problem context. Since these criteria cannot be universally pre-determined, it is up to the analyst to determine them each time she faces a new policy problem.

But what is not relevant for today’s problem may become relevant as the problem context changes or for other problems in the future. So the analyst needs tools that filter data without discarding it (for the long-term) and that are dynamic enough to accommodate changing information (for the short-term). Geva-May proposes the use of ‘checklist, tables, decision trees, and flow charts’ to code and classify political data, the position of actors, and the sequence of problems (1997b). Her proposed ‘flow chart’ is in fact a categorical schema that creates spaces to hold various kinds of policy-relevant information. In its substance and style it looks very much like Werth’s proposed text.

23 “Policymakers also were asked to indicate where they go when they are seeking information to help them with a tough policy issue. It is interesting to note that policymakers did not indicate that they turn to published literature for information at these time-critical junctures. Rather, they said that they turn to people ‘who either know the answer or know where to find it (Sorian & Baugh, 2002).”
world ‘frames.’ The same critiques offered of Werth’s schema also apply here; this method, while useful, is tedious and manual.

Though it is not the ultimate goal of policy analysis, this kind of work (placing chunks of information into pre-existing categories; otherwise called labelling, classifying, coding, annotating, or marking-up) is foundational to subsequent stages of analysis. Nevertheless, in its ideal ‘rationalist’ formulation, the process of identifying the elements of the problem context (events, actors, their values, motivations, goals, constraints, and resources) does not just happen once at the beginning of the analysis, but continually and iteratively throughout the analysis (Geva-May, 1997b; Dunn, 1981c; Dunn, 1981b; John, 1998a).

I propose that policy analysts need a system that retains their analytical and classificational efforts. It needs to allow for iterative incrementation and continuous reclassification of new information in the collection. Classifications that apply in one problem context may not apply later on, but if they are incremented into a structured collection of analyses, it can become a valuable repository of their own work and an excellent jumping off point for future work.

Analysts need tools that will help them understand the intentions and meanings of people from different times and places, with different perspectives, without having direct, first-hand access to these contexts. When reading articles from different disciplines and polls from different counties, analysts need to know what they meant in the original context before they can evaluate its ‘quality’ and ‘relevance’ to the issues at hand. Analysts need tools that help by relating the texts’ authorial and disciplinary vernacular to their own. This is not possible using the current paradigm of trans-disciplinary
'poaching' (where lay people randomly seek bits of information and lacking understanding, still try to piece together their picture of the world patch-work style).

1.5 Traditional Disappointment

While I have spent the better part of four years looking for software to help manage the conceptually complex work of text-dependent knowledge workers, I have not found any robust enough to offer all the necessary functionality in one package, dependable enough to rely on under deadlines, or easy enough to use to merit a broad recommendation.

Apparently today's market lacks the capacity to bring together existing textual analytic technologies in a way that identifies and repurposes the basic facts of a text. While interdisciplinary discovery and argument have substantially refined theories of cognition; and while certain linguists have made great strides towards launching "a genuine science of meaning construction and its dynamics (Jannsen et al., 2000)"; while infrastructural hardware technologies and software strategies have significantly improved the computational capacity of personal computers; policy analysis continues to muddle through, largely unaided by these advances.

My disappointment compounded as I searched among the academic disciplines known for equipping professional domains with technological tools. While I discovered Knowledge Management at the intersection of Management and Organizational Theory and Information and Computing Sciences, I also discovered a doleful track record. Since the history of Knowledge Management is not the proper subject of this thesis, I will try to condense the disappointments of its first two generations briefly, in narrative form.
Knowledge Management is both a consequence and extension of the Taylorian rationalist-empirical impulse to radically redesign business processes based on principles of scientific efficiency. The problem was that these radical redesigns often emptied organizations of the know-how required to achieve business success (Sharkie, 2003). So business process redesigners turned to the ‘problem’ of ‘knowledge’ and began to redesign, or rather, to retrofit enterprises with information systems for decision support (Snowden, 2002). During this first generation, the focus was on the “appropriate structuring and flow of information to decision makers and the computerisation of major business applications (2002).” Redesigners found ready collaborators in the computing and information sciences, and accordingly, the first generation of ‘knowledge management’ tools looked a lot like the data management tools that computer scientists were already working on (McElroy, 1999).

Eventually, around 1995, when people started to realize trying to manage ‘knowledge’ using information systems was missing something fundamental, the redesigners maladapted a distinction between tacit and explicit knowledge that kept busy for a while (Snowden, 2002). Other ideas also surfaced in response to different aspects of the failures of the first generation, including an emphasis on the importance of knowledge production and organizational learning for innovation, the rules of knowledge structure and the process of knowledge lifecycles (McElroy, 1999). New conceptions of ‘knowledge’ replaced previous ambivalence about the term (1999).

But by the late nineties, KM users and theorists realized that systems designed to extract process based information and facts of various kinds from workers were often not well received. The idea of Knowledge Management began to lose currency. To be fair,
certain large enterprises (Xerox, IBM, Dow) have had some striking success in their KM
efforts (Snowden, 2002). Nevertheless, experts agree (Malhotra, 2002; McElroy, 1999)
that knowledge management has “substantially failed to deliver on its promised benefits
(Snowden, 2002).”

I would add two explanations for this general failure. First, early efforts lacked a
coherent epistemological basis. Initially, knowledge was not well-defined. Often it
simply meant a nominally higher order of information or data. Subsequently, they
mistakenly thought that knowledge could be ‘managed’ – packaged and delivered in a
controlled environment. But contrary to the fondest hopes of modern KM endeavours
like the Semantic Web (Ludwig, O’Sullivan, & Zhou, 2004), extra data about information
is not knowledge.

Of course the difficulty with philosophical questions is in not getting lost in issues
lurking beneath the question you ask. This is especially true in this case, since questions
of epistemology are closely coupled with those of ontology (McElroy, 1999). Apparently,
the first generation found it expeditious to ignore these questions altogether (McElroy,
1999). The default position of the second generation seems to favour adopting a Kantian
epistemology, which viewed knowledge “as a thing, something absolute, awaiting
discovery through scientific investigation (Snowden, 2002).” But knowledge is not
simply a thing; it is more of a process. And this process, while it can be used to facilitate
group awareness, is nevertheless inherently a product of individual cognition.

The second explanation of knowledge management’s failure to date has to do with
a focus on organizational value propositions while ignoring the individual value
proposition. The idea here is that without a suitable motivation to use a tool, most people
just won’t use it. If they see it as just another step that they are expected to take but one that they can skip, they likely will. In order to create a motivation towards adoption, system designers must prove an immediate value to the individual user (Malhotra et al., 2005).

In 2002, Snowden argued that if only the academics and management involved in the field could make the significant conceptual change required, “effectively bounding or restricting over a hundred years of management science,” then Knowledge Management could enter its third generation. By way of a historical example, Snowden envisions the magnitude of this change to “the phase shift from the domination of dogma in the late medieval period, to the enlightenment – moving from esoteric complication to a new simplicity based on a new understanding of the nature of meaning (2002).”

The third generation requires the clear separation of context, narrative and content management and challenges the orthodoxy of scientific management. Complex adaptive systems theory is used to create a sense-making model that utilises self-organising capabilities of the informal communities and identifies a natural flow model of knowledge creation, disruption and utilisation (2002).

Snowden’s sense-making model draws out a key difference between natural and human complex systems.24 Scientific management made the mistake of assuming that the imposed order of human complex systems is an absolute or universal structure, as in nature (2002). Snowden points out that the stability and usefulness of human complex systems is based on “common will and a stable environment (2002).”

When conditions of uncertainty are reached, the order can break down or artificially persist beyond its usefulness. By implication it is argued that the dogma of scientific

24 A complex system comprises many interacting agents, an agent being anything that has identity (2002).
management, hypothesis based consulting and the generalisation of best practice from multi-client or multi project studies, are inhibiting factors in progressing to the new levels of conceptual understanding required in the modern world (2002).

The idea is that the (post 9-11) world is highly uncertain. While critics of rationalist and empiricist traditions might argue that the ‘certainty’ of prior eras and mindsets was only ever an illusion created by dogmatic adherence to comforting assumptions, certain historic events have a way of challenging and changing assumptions. Older models of management and decision making are calibrated for certainty and cannot be expected to provide the conceptual or organizational leadership required by current conditions. Snowden’s model helps group participants to use their own self-organizing capacities to become aware of how they see things, to break down non-useful perspectives and build new ones. Snowden advocates using his model to enable ‘descriptive self awareness’ within an organisation “rather than imposing an pseudo-analytic model of best practice, it provides a new simplicity, without being simplistic, enabling the emergence of new meaning … in a complex ecology of knowledge (2002).”

Snowden’s model assumes that the context of human knowledge includes two dimensions; abstraction and culture. Using three examples of knowledge sharing, he illustrates that the appropriate level of abstraction for knowledge sharing is determined by the degree to which speakers share context. Snowden adapts a single quadrant I-Space model (Boisot, 1994) that contrasts the level of abstraction of knowledge with the cost of

---

25 In the first, knowledge is quickly transferred between trusted co-workers. In the second, when an unknown colleague asks the expert a question, a conditional response is offered only after lengthy discussion. In the third, the expert is asked to codify their knowledge. If interested, they write a book.
its codification and identifies the theoretical upper and lower bounds for acceptable levels of abstraction (See Figure 1).

High abstraction either involves expert language, taught in universities, through books, training programmes and so on, or shared experiential and cultural referents. At the highest level of abstraction, where I share knowledge with myself there is a minor cost, I may keep notes but no one else has to read them. On the other hand if I want to share with everyone the cost becomes infinite, as the audience not only need to share the same language, but also the same education, experience, values and so on. In practice there is a very narrow zone between the lower and upper levels of acceptable abstraction in any knowledge exchange. (Snowden, 2002)

Culture, the other dimension of context, is defined somewhat schizophrenically as both ‘the observable results or artefacts of social behaviour,’ and ‘the underlying system of shared beliefs, concepts, rules and meanings expressed in the ways that humans live.’ Snowden maintains that “both cultures are key to the flow of knowledge within an organisation (2002).” He frames the transfer of knowledge as a cultural imperative that takes the form of learning and of teaching. Whereas teaching relies on the culture of stability and certainty, learning requires the culture of fluidity and ambiguity.
Snowden’s sense-making model (the ‘Cynefin’ model is named for a Welsh word for situatedness) is a post-empiricist variation of a quadrant matrix. It “creates four open spaces or domains of knowledge all of which have validity within different contexts. They are domains not quadrants as they create boundaries within a centre of focus, but they do not pretend to fully encompass all possibilities (Snowden, 2002).” Nevertheless, the model is intuitively comprehensible to those familiar with matrices. The vertical border between domains represents the contextual dimension of Abstraction and is labelled at the extremes (High and Low). The horizontal border represents the Cultural dimension of context and its extremes are labelled Teaching and Learning.
The pair-wise properties described by each Quadrant (High Abstraction, Teaching; Low Abstraction, Learning) are used to describe different domains of knowledge and their ideal KM strategies. These domains also map onto the three kinds of systems described in Systems Theory: complicated, complex, chaotic. Complicated systems are further separated into “those in which we know all of the cause and effect relationships and those that are knowable if we had the resource, capability and time” (2002). Each of the domains contains a different model of community behaviour – and each requires a different form of management and a different leadership style (2002).
The purpose of this model is to allow groups to share the experience and see the implications of knowledge flow across these boundaries. While many flows are possible, Snowden describes in detail “an idealised model of knowledge flow involving three key boundary transitions – the disruption of entrained thinking, the creation and stimulation of informal communities and the just in time transfer of knowledge from informal to formal (Snowden, 2002).” These ‘key transitions’ are shown in Figure 4.
In Figure 4, traditional perceptions about what is ‘knowable’ are disrupted by a descent into chaos. The realization that entrained thinking no longer works in the context of chaos forces a different, more active kind of sense-making that throws people together into temporary communities. They must work together to sense and respond to patterns in the chaos. Pattern recognition crosses the border into complexity, where ways to manage the patterns are identified. When causes and effects can once more be known, the flow of knowledge returns to the knowable, where, with the necessary investment of time and money, the knowable can become ‘known.’ Snowden maintains that the value of this model is in group explorations of the dynamics of these boundary transitions in the context of the organization and from the perspective of its members.

While Snowden’s ideal of a third generation of knowledge management is an exciting departure from its predecessors, it is not clear that these ideas have taken hold, or that we have actually entered this third generation. This model is exciting to me, because
it explicitly shifts the focus to the context, narrative, and content of discourse. It goes further, emphasising the role of cognition in the creation of meaning. Most importantly the model is predicated on the innate human capacities to impose patterns in chaos; to recognise, disrupt, reinforce and seed the emergence of patterns in complexity; to collectively impose order through accepted practices to create predictable environments; and to develop and disrupt the perceived wisdom of expertise.

While the ‘personal knowledge management’ (PKM) that I propose is drawn in counterpoint to much of the existing domain of Knowledge Management, it would certainly fit within the rubric of Snowden’s third generation. But rather than focusing on the development shared organizational knowledge, my work emphasises the organization of experiential personal knowledge. In PKM, many of the technical contributions of the first and second generations are still important (meta-data extraction and management, document imaging and management, single-source content management, semantic tagging and XML), though these processes should be automated. What is centrally important, however, is the concept of managing the content and process of your own learning.

For example, in PKM, effective file management is indispensable. File naming conventions, directory structures, processing priorities, automating batch processes, file conversions are all pedantic non-problems to anyone who has thought through the issues and created a working system. But many haven’t. And most will never take the time necessary to do so for themselves. If it must be done, (and it must) it should be done automatically by the computer, not the user.

So hard drives have become one of the hardest places to find what you need. I offer the messy hard drive as a metaphor for the organic personal knowledge base. Chaotic systems exhibit a certain quality of dependability. The mess is a more or less stable situation, it just can’t usually be expected to produce anything else but more mess. In order to take that personal knowledge base and make it productive and useful, I will have to find a way to reliably call previous knowledge into play as current circumstances require.

This highlights a key assumption of the model; that it would be best to draw first from the existing knowledge base. One reason for this supposition is that ‘old knowledge’ has presumably been vetted (subconsciously categorized, for example, as believable, hocus pocus, bad science, political doublespeak, etc.) and if it could be mobilized as such, it should prove more valuable as a starting point, than any wide swath of raw information that a library or web search would be likely to produce. New information must be read. Categorized. Understood. And this costs time and money.
Rather than theoretically trying to organize and index the full range of human experience, PKM takes interactions with texts as proxies for mental processes more generally. It uses digital texts, since they are digitally manipulable. But trying to organize an incomplete and largely unrecorded trajectory through a lifetime (or even a few years' worth) of digital text is next to impossible using current off-the-shelf software (and even testing a few current open-source and experimental packages). So I am trying my best with the best software I can find. Along the way I am drawing up a list of deficiencies to help in calibrating my vision of an ideal set of software tools for this job.

Rather than just find a way to mark up a text with my idiosyncratic understandings and perspectives; I want to gather my views in a way that automatically situates them with reference to a universally applicable system of reference: an upper ontology that can accommodate any structured knowledge representation from any discipline. In Snowden's terms, taking my notes in a system that effectively maps between the text, my interpretations, and your interpretations will reduce the transaction cost significantly.

So, before I read a text, it should be pre-processed so that the evolution of my 'subjective' interpretations of texts can be evaluated compared to the many 'objective' dimensions of deep structure in a text that have been automatically 'surfaced' for me. I can keep this multi-dimensional complex invisible as I read, or audit and explore any of these dimensions structure at my leisure. Then I can reliably hang my qualitative analysis on a firm, 'objective,' or quantitative foundation. I can share my readings of a text with you; and compare yours with mine. We can learn from each other with direct reference to what (and how) we already 'know.'
But, meanwhile, until the ideal vision is realized, what tools can a Policy Analyst use to make better use of what they ‘know’?
CHAPTER 2:
WHAT ARE THE AVAILABLE TOOLS GOOD FOR?

2.1 Operational Assumptions

My thesis regarding the usefulness of knowledge management tools to policy analysts rests upon several assumptions. First, technologies must be able to prove an immediate value to the individual user in order to create a motivation towards adoption (Gill, 1996). An associated premise here is that motivated adoption of 'smart practices' (Bardach, 2000a) will lead to better analysis (Geva-May & Wildavsky, 1997). However, in order to merit 'motivated adoption,' smart practices ultimately need to prove useful, usable, and well adapted to the context of their intended use. Initially, however, users must somehow discover the tools. Once they are found, they must be evaluated. Once evaluated, they must be adapted, qualified, or rejected; and then recommended, as appropriate (Van den Bulte & Lilien, 2002).

I argue that existing knowledge management tools for textual analysis will require extensive adaptation before they will merit motivated adoption by policy analysts and other knowledge workers.27 The novel contribution of this thesis turns on the generation of recommendations for improving existing tools. In particular, in Chapters 3 and 4, I work at the development of the rationale for, and partial implementation of an interdisciplinary analytical rubric that aims to facilitate automated evaluation and analysis of interdisciplinary collections of text. Initially, this thesis will have implications for the

---

27 Notwithstanding failure to meet the needs of Policy Analysts, after over 4 years of product testing and evaluation, I can provide a qualified recommendation of a wide range of tools and their suitability to particular tasks that familiar to many knowledge workers. See Section 3.5.1 through 3.5.10.
policy analyst and individual knowledge worker that could impact their daily computer-based work routines and capacities. Ultimately this work may also have upstream organizational implications, as efficiencies captured by individuals have a positive impact on the organization’s overall success.

In this Chapter, I evaluate and recommend textual analysis tools and personal knowledge management strategies tailored to the policy analyst. In Section 2.2 Demanding Environments, I discuss the working environment of policy analysts as a basis for deriving Evaluation Criteria in Section 2.3. In Section 2.4 Technological Tools, I describe how software tools are classified according to their intended user base, by their functionality, and of particular importance to analysts on a tight schedule, by their level of automation. In Section 2.4.2 Hardware Description, I outline the technology infrastructure required to use and evaluate these software tools. Section 2.5 Suitability Evaluation assesses ten classes of textual analysis or knowledge management software, identifies top ranked programs, and discusses the degree to which a policy analyst may find the tools useful. In Section 2.6 Tailored Solutions, I describe the ideal functionality of textual analysis and personal knowledge management software for policy analysts, provide some examples of ‘smart practices’ I have developed to manage a large transdisciplinary collection (See Appendix D Smart Practices), and explain how the ideal automated classification software requires a classification schema that can function regardless of the disciplinary origins of a text.

2.2 Demanding Environments

Analysts work in a demanding, time-sensitive environment. Policy analysts are professionals who are trained to do the best they can with the time and information they
have available (John, 1998b). They usually have little enough of either (Geva-May, 
1997b). They are trained to grapple with the issues facing decision makers and wrestle 
them into manageable policy problems. Even without full or adequate information 
(Simon, 1957) they are taught how to identify, extract and synthesize competing 
priorities, values, and perspectives surrounding a policy problem and how to turn them 
into evaluative criteria (Dunn, 1981c). They learn how to come up with various courses 
of action which could satisfy these priorities and ultimately how to use their evaluative 
criteria to compare, rank and recommend policy options (Kingdon, 1995b).

Policy analysis depends, at least in part, on the information contained in 
documents (Bardach, 2000b). The variety of documentary sources used by policy 
analysts is discussed in Section 1.4 Intellectual Motivation (on p. 24). However, 
documented information is not the only policy-relevant resource for analysts. Formal and 
informal social networks serve as vital channels of information (Schneider, 1992). For 
example, in Kingdons’ (1995c) account of policy windows, information gathered through 
social channels is invaluable to effective brokering and manoeuvring in the halls of 
power.28

Most executive and bureaucratic systems have made the transition, to varying 
degrees, from legacy paper to mainframes and then to networked workstation systems 
(Snowden, 2002). But despite the ubiquitous presence of desktop computers in both 
government and the private sector paper still dominates offices today (Lyman & Varian, 

---

28 This kind of knowledge is unavailable from any other source; it is ephemeral and time-sensitive. It is an 
excellent example of what the discipline of Knowledge Management calls 'tacit knowledge.' (Von Krogh, 
Ichijjo, & Nonaka, 2000) It is uncodified knowledge in that it does not exist on paper or in digital form. 
The theory holds that tacit knowledge is embedded in the social fabric of an organization, though it 
ultimately resides in the minds of the individuals who make up an organization.
In any case, not all policy-relevant information is in the ideal digital format that is readily commensurable to textual analysis using existing KM tools. And since most available textual analysis tools are particular about the type of document they can handle (Lewins & Silver, 2006), even having a document in digital format is no guarantee that the tools can process them. For these practical reasons I have made a concerted effort to evaluate tools that convert both into and between digital formats.

While not all policy relevant information is documented, and not all documents are analysable using textual analytical software, the argument for helping analysts manage their policy-relevant textual information more effectively remains convincing given the volumes of textual information involved in the analysts’ daily routine (Sorian et al., 2002; Gross, 1971). Yet analysts cannot be expected to use tools that they don’t know about, or that they lack the resources to use. Somehow they must manage to find and finance the tools that are right for the job. As citizens and taxpayers, the public has a reasonable expectation that policy analysts will not use tools that do not offer a clear advantage in their work. So analysts must somehow come to know whether or not any

29 Simply asking analysts what digital tools they use is problematic, because Sorian (2002) suggests that analysts may not be using any. So how can we explain non-use of potentially helpful tools? One reason may be that people are unlikely to select tools they are not personally familiar with. Ignorance can be explained vis-à-vis the high cost of acquiring knowledge.

The cost of evaluating tools is high. Finding tools that can streamline everyday work and facilitate specialized analysis for our ‘case study’ knowledge worker has taken significant, unfunded amounts of personal time and money to integrate the tools into a functioning computer system. Inevitably, many tools have failed to meet the criteria of ease-of-use, dependability, customizability, or cost. The cost of finding failures can significantly wipe out the perceived gains of any successes tools.

However, not knowing what tools exist also costs the analyst in time spent doing things in ad-hoc, low-yield, analog (paper or ‘pile’ based) ways. The cost of using paper-based systems is not just the time spent wrangling paper and the money spent feeding the photocopier. If greater efficiencies are possible (lower seek-to-find time ratios) using digital resources and tools, efficiencies can be expected to compound with continued usage (in a system that scales well). So delayed adoption compounds its costs in lost efficiency.

30 While one class of KM tools exist that can facilitate the conversion from paper to digital documents, as described in Section 3.5.1 Optical Character Recognition, the conversion process is error-prone and time consuming, and not likely to be a part of a policy analysts’ daily routine. Another class of KM tool transcribes spoken language into digital text (See Section 3.5.2 Voice Recognition) and potentially offers more value to the policy analyst as a keyboard replacement.
‘added value’ exists for a given tool; ideally without incurring the high costs of evaluation themselves.

That is where this thesis can be of practical value to policy analysts. The ‘tech savvy’ elite may know of numerous online resources for general technology reviews (such as cnet.com, gizmodo.com, arstechnica.com, pcmag.net, macworld.com, theregister.co.uk), but finding an evaluation of software tools targeting your chosen profession is a rare find.

### 2.3 Evaluation Criteria

This Evaluation Matrix can be used by a Policy Analyst wanting to evaluate other software. First, fill in criteria values. Next, rank options, as in Goeller scorecard method (See Appendix B, Example Policy Brief for a completed example) with color, using green and red to highlight best and worst options.
Criteria for evaluating the suitability of the various software tools identified are based on cultural and organizational circumstances of the policy analyst’s daily work. Given the expectation of tight turn-around times for policy briefs and memos, tools must not require too much time to set up or complete analysis, where ‘too much’ is any duration above a set ratio of available time to average time-required-for-analysis. While not essential, near instantaneous real-time analysis or retrieval of stored, pre-processed analyses are preferred.

Given the cost constraints facing most organizations, the monetary purchase price and maintenance costs must be minimal or within reach for the individual, since organizational finances may not be available. While not essential, free is preferable. The associated time costs of maintenance must be either low or reasonable, where ‘reasonable’ is any duration below a set ratio of productive time to maintenance time.

Two criteria are extrapolated from software usability’ studies (Hartson, Andre, & Williges, 2003). The tools’ level of complexity is rated in terms of their perceived complexity from 1 to 10; where 10 is most complex. The perceived complexity of tools must not exceed a set threshold. Tools must be dependable, where dependability is a rating above a set ratio of program errors or failures per hour of productive use. If the tool is too complex or too unstable, it should not be used.

Finally, given its policy context, this evaluation includes two criteria relating to the institutional or political feasibility of implementing the proposal. Tools must be rated in terms of their perceived usefulness from 1 to 10; where 10 is most useful. For example, the preferred format for source documents is XML, PDF, DOC, RTF, TXT, in that order.
A tool that only accepted TXT files as its input file format would generally not rank as high, since it is not universally usable with digital texts of any format. Also, tools must be used often, where ‘often’ is a rating above a set ratio of times-used per week over total possible-times-used per week. If an alternate method of accomplishing a task is preferred or if its perceived usefulness is too low, it should not be used.

2.4 Technological Tools

2.4.1 Software Classification

The software industry generally produces software with several types of end-user in mind. Some are consumer grade products, intended for the average home-based computer user. Others professional grade, intended for a smaller set of users whose professional activities can justify higher price of the software as a reasonable cost of doing business—so long as the software delivers upon its promised functionality. Still others are ‘enterprise level solutions.’ Their price points and breadth of functionality are generally so great that only large enterprises (corporations, certain non-profits, and governments) could ever need or justify the expense.

In the nineties, software tools for managing organizational digital information flows were primarily offered at the enterprise level. These included portals to data warehouses, enterprise-wide messaging systems, groupware (i.e. Lotus Notes), content-management systems for digital publishing, and various other applications intended to gather ‘metric-able’ data from workers or consumers (McElroy, 1999). At the professional level, offerings included business intelligence tools like contact-management databases and ‘personal information management’ (or PIM) software bundled with
handheld organizer devices that included Calendars, Address books, and Task scheduling that could be synched with more robust Desktop versions of PIM software, such as Microsoft's Outlook®. At the consumer level, scaled-down PIMs looked like digital address books, Email clients and web browsers.

For this thesis most of the evaluated software packages were professional grade tools, with the exception of some general-purpose consumer-grade productivity tools. The software of primary importance to this project includes professional grade tools that an analyst may not already be familiar with. This excludes Web browsers, email clients, and business productivity tools like the Microsoft® Office Suite. While enterprise tools certainly exist in the software categories I evaluated, I could not afford to secure them for testing. Wherever possible, I used free evaluation software, and when the software seemed to merit further testing, I purchased a student licence.

The market further classifies existing software tools based on their functionality, which Tyndale (2002) has formalized in a taxonomy of Knowledge Management Tools (Earl, 2001). This thesis evaluates ten types of software that offer a range of KM-inspired functionality. Certain other types of software that could have been evaluated (Data Mining, Desktop Search, Desktop Macro Scripting, Batch File Utilities, Time Management, Note Taking, Mind Mapping and Brainstorming) were not formally evaluated due to time and financial constraints. Nevertheless, these 'unevaluated' software classes offer important functionality that cannot be ignored when crafting an ideal picture of the 'ideal tool set' for policy analysts. (For more on this, see Section 2.6 Tailored Solutions.) Also, a variety of digital document conversion utilities were identified only as needed to facilitate the evaluation of the primary software packages.
These, too, were not evaluated with the primary software packages. Many of these unevaluated tools were nevertheless used extensively throughout the course of this thesis.

Finally, existing KM and text analytical tools fall into one of the following four categories: 1. Manual analysis; 2. Software-aided manual analysis; 3. Semi-automated analysis (using pre-existing coding strategies), and 4. Automated analysis. *Across all classes of software, automated analysis is preferred.*

### 2.4.2 Hardware Description

The hardware used in the course of this thesis includes both the latest and some very old technologies. The computer on which I ran all of my software evaluation and attempted textual analyses is an HP Compaq TC1100 Tablet PC, running Windows XP Professional Tablet Edition. This ultra-portable laptop has a 1.2 GHz Intel Pentium M Centrino processor, 1.5 GB of RAM, and a 60 GB hard drive. Besides the traditional keyboard and mouse, both of which are detachable, the Tablet PC has a pen-based touch screen. The Tablet features embedded handwriting and voice recognition for when the keyboard is removed. I also used an HP ScanJet 4600 as a scanner, a network-attached Canon photocopier, and a very old, very slow 533 MHz Pentium III as a backup PC. Since I do not own a Mac, several gracious friends and family permitted me to evaluate certain tools on their systems.

---

31 Over the course of this thesis, I had to return the HP scanner for replacement nine times, and the HP computer for repairs eight times. HP eventually returned my money for the scanner. The original computer had everything in it replaced at least twice before HP sent me a brand new upgraded replacement. Since the replacement has also failed repeatedly and they no longer manufacture the TC1100, HP is in the process of sending me a different top-of-the-line tablet style laptop to replace it. In some senses, my experience can be read as a cautionary tale for the technology enthusiast: the latest is not necessarily the greatest. Despite my rocky experience with HP hardware, and though I am not likely to purchase another HP product, I will certainly purchase another tablet. The ability to integrate pen and digitize handwriting is excellent for note-taking and free-form sketching. There are also a range of pen-enabled productivity software that make the tablet a must-have in portable computing.

32 The scanner and photocopier were used to evaluate OCR software (See Section 3.5.1).
Since beginning this project, I have realized that what I need to complete the analyses I have attempted is a high-powered, server-class desktop computer, or access to a supercomputing grid.\(^{33}\) In any case, trying to create an automatic document classificational schema and a semantic network using all the words in 2,500 documents has exceeded the capacity of each system I have tried to run the analysis on to date.

The idea of one computer that can effectively turn all manner of human symbolic expression into searchable digital documents has motivated me to put the idea to the test. And despite being beleaguered by hardware and software troubles, I am nevertheless impressed with the basic pen-based interface and handwriting recognition. The tablet is a valuable hardware innovation that makes a range of new software tools possible.

The ability to take a range of previously 'tacit' or otherwise inaccessible knowledge and convert them to digital texts only serves to highlight the need for thorough and expandable textual analysis tools. The explosive growth of various file types calls for an intuitive file management system that can automatically file documents, find and group duplicate versions in different formats, extract or retrieve document and bibliographic meta-data, auto-classify documents based on contents, provide intelligent summaries, highlight and foreground the most important relationships between concepts, and offer a seamless and sharable digital margins in which to annotate texts. This is admittedly a high standard; and one to which I do not expect current software to attain.

\(^{33}\) However, none of the programs evaluated are multi-threaded applications, nor are they 64bit compliant, so running them with the latest hardware may not resolve the inherent limitations of 32bit applications.
2.5 Suitability Evaluation

The method used for this evaluation is drawn from policy analysis. (For further details, See Appendix B, *Example Policy Brief.*) In the following sections, for each class of software tools, the name of the evaluated software, the top-ranked tools, a brief discussion including a list of shortcomings, and a brief assessment of the tools suitability are provided.

<table>
<thead>
<tr>
<th>Classes of Software Tools</th>
<th>Functionality Offered</th>
</tr>
</thead>
<tbody>
<tr>
<td>*File Conversion</td>
<td>Converts between digital file formats.</td>
</tr>
<tr>
<td>Optical Character Recognition (OCR)</td>
<td>Converts images of analog paper documents into editable digital texts.</td>
</tr>
<tr>
<td>Voice Recognition</td>
<td>Converts analog spoken words into editable digital texts (or computer commands).</td>
</tr>
<tr>
<td>Handwriting Recognition</td>
<td>Converts written words into editable digital texts.</td>
</tr>
<tr>
<td><strong>Meta-data Management</strong></td>
<td>Structures &amp; stores data about digital files.</td>
</tr>
<tr>
<td><strong>Bibliographic Reference Management</strong></td>
<td>Structures &amp; stores data about published documents, &amp; mobilizes data for citation while writing.</td>
</tr>
<tr>
<td><strong>Document Management</strong></td>
<td>Names &amp; locates digital files.</td>
</tr>
<tr>
<td>Content Management</td>
<td>Separates a document's content from its formatting &amp; presentation to allow centralized management and reuse of textual content.</td>
</tr>
<tr>
<td><strong>Content Analysis</strong></td>
<td>Enumerates frequency of terms &amp; concepts in digital texts.</td>
</tr>
<tr>
<td>Qualitative Data Analysis (QDA)</td>
<td>Facilitates classification (coding) &amp; modeling of concepts in digital texts &amp; AV files.</td>
</tr>
<tr>
<td><strong>Semantic Network Analysis</strong></td>
<td>Identifies &amp; measures relations among concepts in digital texts.</td>
</tr>
<tr>
<td><strong>Data Mining</strong></td>
<td>Searches databases (structured data) for previously unknown patterns.</td>
</tr>
<tr>
<td><strong>Natural Language Processing</strong></td>
<td>Attempts to create an human-computer interface using human language as humans use it.</td>
</tr>
<tr>
<td><strong>Parsing and Chunking</strong></td>
<td>Identifies and parses sections, paragraphs, sentences, phrases, words and punctuation in digital texts.</td>
</tr>
<tr>
<td><strong>POS-tagging</strong></td>
<td>Identifies parts of speech of words in every sentence in digital texts.</td>
</tr>
<tr>
<td><strong>Morphological Analysis</strong></td>
<td>Identifies lexicographical word roots in digital texts.</td>
</tr>
<tr>
<td><strong>Word Sense Disambiguation</strong></td>
<td>Attempts to disambiguate possible word senses.</td>
</tr>
<tr>
<td><strong>Word Stemming</strong></td>
<td>Identifies lexicographical word roots in digital texts.</td>
</tr>
<tr>
<td><strong>Lemmatization</strong></td>
<td>Removes variant endings and reduces words to their lexicographical word roots.</td>
</tr>
<tr>
<td><strong>Document Structure Tagging</strong></td>
<td>Assesses the structure, format, layout, and embedded objects of a text to identify and extract its Chapter Titles, Sections, Authors, Dates, Page Numbers, Images, Diagrams, Equations, Charts, and Tables and then automatically normalize their formatting.</td>
</tr>
<tr>
<td><strong>Semantic Mark-up</strong></td>
<td>Identifies function or role of portions of digital texts.</td>
</tr>
<tr>
<td><strong>Automated Entity Recognition</strong></td>
<td>Recognizes &amp; extracts a variety of 'entities' from a text.</td>
</tr>
<tr>
<td><strong>Content Summarization</strong></td>
<td>Summarizes most important features of digital texts in natural (plain) language.</td>
</tr>
<tr>
<td>*Desktop Search</td>
<td>Locates digital files on local storage media (hard drives).</td>
</tr>
<tr>
<td>*Time Management</td>
<td>Facilitates organization of events and priorities.</td>
</tr>
<tr>
<td>*Note Taking</td>
<td>Facilitates recording events through handwritten or recorded AV notes.</td>
</tr>
</tbody>
</table>
2.5.1 Optical Character Recognition (OCR)

*Evaluated:* OmniPage Pro 14 Office; OmniPage Pro 15; ReadIris Pro 8; ReadIris Pro 10

*Top Ranked:* OmniPage Pro 15, ReadIris Pro 8 (free)

Document imaging was one of the original KM tools. It has come a long way since the early nineties. Enterprise level OCR products marketed today have the ability to read barcodes and hand written forms. These features do not add value for the policy analyst. Over the years, the algorithms for matching characters have been extended with lexical tools like probability based word matching from dictionaries and corpus statistics. These variants are called Intelligent Character Recognition (ICR). Professional grade products offer the capacity to customize and automate workflows so that the system is continually processing documents or watching folders for documents to process.

While their ability to replicate the formatting of the original document is usually impressive, in documents with complex structures, such as alternating multi-column layouts, their success rates drop. Also, OmniPage has trouble determining the font case (normal, bold, italic) of serif fonts, and will occasionally alternate font families (Times New Roman, Arial) from page to page. Having said that, scanning documents above the recommended 300 DPI seems to resolve the font case issues.
Recognition of Diagrams, Images, and Charts is very poor. Recognition of mathematical equations is not supported. Batch automation calls cannot be throttled and often crash. User intervention is required at numerous intervals; to check the zoning of the document; to train the recognition of characters; to check the recognition of words; to spell check the document; and to check the formatting of documents. Since all errors require manual manipulation, the OCR process is highly time intensive. Creating workflows to minimize user intervention generally produces unusable documents, so if the document conversion is worth doing at all, it is likely worth spending the time making the necessary corrections.

While policy analyst will not generally be expected to convert organizational documents into digital format, on a practical, day-to-day level, certain paper-based resources may be more valuable if they are digital and searchable. Most scanners and many photocopiers now offer the option to scan or copy straight to PDF, and those with access to Adobe Acrobat Professional can make use of built-in OCR technology to make all scanned PDFs searchable. However, better results can be obtained using OmniPage. Only those analysts most committed to building a resource record for themselves will scan regularly. In some cases, OCR can be used to convert a document into a more manageable (RTF or DOC) file format with more reliability and quality than a native ‘Save As’ function, for example, in Adobe Acrobat Professional.

2.5.2 Voice Recognition

*Evaluated:* IBM ViaVoice 5; Dragon Naturally Speaking Professional 14 and 15

*Top Ranked:* Dragon Naturally Speaking 15
Voice Recognition Software is finally usable. This software can be used to dictate any text into emails, documents, spreadsheets, web browsers, and to control applications on your computer. While previous versions required users to spend some time training the software to their voice, newer versions work very well out of the box. New versions work with Bluetooth (short-range wireless) headset mics. Analysts may find this software useful for dictating notes, emails, or reports.

2.5.3 Bibliographic Reference Management

*Evaluated:* Reference Manager 11; RefWorks Student; RefWorks Annual; EndNote

*Top Ranked:* Reference Manager (PC), RefWorks (online), EndNote (Mac)

This class of software manages bibliographic information of your collected published documents. It also allows users to insert and automatically format and manage citations in reports or documents written in Word. This functionality is excellent for heavy writers.

However, RefMan has a non-standard user interface, so navigating between database fields is not simple. Also, the software is rarely updated. While RefWorks shows innovation and is a dependable product, it requires an active internet connection so is not ideal for all users. Policy analysts who compile reports from previously published materials that require accurate citation of referenced works will likely find this software a great help when writing reports. On the other hand, maintaining an up-to-date citation database can be time-consuming, and for those who do not regularly cite materials, updating and managing the database may prove to be more trouble than it is worth.
2.5.4 Qualitative Data Analysis (QDA)

Evaluated: Atlas.TI; QSR NVivo 4; QSR NVivo 7; Provalis QDAMiner; Qualrus; MaxQDA
Top Ranked: QSR NVivo 7

This class of software is used to manually mark up the contents of text-based document in a structured and searchable system (Lewins & Silver, 2005; 2006). Before the advent of the computerized workplace, this work was done by academic researchers by hand using paper systems. Currently, textual analysis tools can be divided into three categories along the dimension of automation. QDA software generally falls into the category of ‘software-aided manual analysis.’ Ideally, I had hoped to find one that offered fully automated analysis. Unfortunately for the purposes of my thesis, these products are marketed primarily to academics who want to be ‘at the helm’ when their textual collections are ‘coded’ (classified according to a pre-existing or evolving ‘coding’ schema).

Atlas.TI is the only package to offer a ‘semi-automated analysis;’ but running it takes time and slows your computer significantly. Its ‘find all’ (semi-automated coding) search function must be run individually for each code in the coding schema. If new codes or documents are added, the semi-automated coding must be run again for each code. The semi-automated coding ATLAS.ti offers is therefore suited to stable collections (no adding more documents half-way through a project) with very limited, small coding schemas.

While NVivo 7 does not offer automated coding in the sense that I am looking for, if your documents are all formatted identically (using MS Word’s hierarchical Heading Styles) you can automatically code a document by its format. (So every
paragraph under the formatting style ‘Header 1’ is assigned a given code, while text following ‘Header 2’ is assigned another code, and so on.) This functionality is excellent for those generating new documents, as from focus-group or interview transcripts, which can be formatted uniformly. In practice, however, this approach to automated coding fails when coding a collection of documents from various ‘real world’ sources, or whenever the user cannot determine the formatting of their source documents.

NVivo 7 is not as dependable as Atlas or QDAMiner, (it crashed in 8 of 10 test runs). However, it offers more functionality for visualizing and modelling relationships between codes. Also, it is built on a core Microsoft database technology to ‘look and feel’ just like MS Office Outlook, so users familiar with the Microsoft user interface will feel right at home.

The best feature about QDA software is the ability to classify any stretch of text using any category; either with one that already exists, or by creating a new one using the highlighted text for the name of the new category. This allows its users to capitalize on their innate associative and classificational capacities that can relate totally discrete words or ideas as instances of a single category (like happy and sunny). Also, users are not ‘bound’ to a single category scheme. They can create new categories as needed, as they read through, or re-read a text at different points in time.

However, this same feature can be a drawback when two categories exist for the same thing, especially when their names are not nearly identical. In most packages it is relatively simple to merge duplicate or related categories, but this is just another step required to keep the classificational scheme ‘pruned.’ This can be a significant job for large (100+ codes) categorization schemes.
Generally, if this class of software could be configured to run automated analyses with pre-configured yet expandable classificational schemas, this class of software could prove a very useful to policy analysts. Until that point, however, the time involved to set up a project, create a classificational scheme, and run a coding analysis is more suited to longer, academic time-scales. Policy analysts may find the current incarnation of these tool of use on longer term evaluation projects.

2.5.5 Document and Meta-data Management

*Evaluated:* iTunes; PaperPort 10; PaperPort Pro 11; Mercury Document System; Columbus Personal; PDF Explorer; Zoot; Adobe Acrobat Pro 7 Organizer; DevonThink Pro Office (Mac)

*Top Ranked:* iTunes; DevonThink Pro Office (Mac); PaperPort Pro 11 (scanning); Mercury Document System (collecting); PDF Explorer (editing meta-data)

Together with document imaging (OCR), document management is one of the oldest KM tools on the market. The purpose of this class of software is to organize digital documents. While most computer-users may do this by hand, the principle is that the manual approach is not suited to large organizations or large collections. In its ideal form, the process of document management is completely automated and invisible to the user. However, the only program evaluated that offers this level of automation (including naming the directory structure and relocating files as necessary) is the popular consumer-grade music collection software, iTunes. Of course, a well managed music collection will not help the policy analyst as much as one might hope.

How does iTunes do what it does so well? The answer partly rests on its use of document meta-data. Most popular music file formats come with specified meta-data slots, where details about the music file can be entered in plain text. iTunes makes use of this embedded document meta-data to automatically organize and dynamically present a
users’ music collection in an easy-to use, dependable, and elegant manner. When users want to turn their CDs into MP3 files, iTunes reads the CD barcode, checks an online database for the album track information, and automatically enters the meta-data in the appropriate fields. For files missing this meta-data, users have the opportunity to fill in the meta-data fields at their leisure.

Lessons learned from the successful management of music files should be applied to other digital textual and multimedia files. In the real world, documents are split into sections, some of which can stand alone (Books have Chapters, Journals have Articles, Websites have Sections, and all of the above have pages.). This means that some documents have implicit relations with others (just as certain songs belong to a given album) that should be recognized and operationalized.

In fact, all digital file formats (like .DOC, .PDF etc.) have embedded document meta-data slots. In practice, however, it seems as though most people do not bother with entering this information. So most digital texts have spotty, incomplete, or empty meta-data. And since meta-data fields accept just about any plain text strings, their contents are not dependably uniform between documents.

Unreliable document meta-data is one reason that textual document management tools do not offer the same level of automation as their music file counterparts. At best, they are an alternative to throwing all your files in one folder, or creating a labyrinthine directory structure to accommodate all your files. They can help a user to organize their files with a different interface than Windows Explorer, but they do not take the responsibility entirely off of the users’ hands. Most do so by offering the user a nominally
easier way to enter document meta-data, by auto-naming new files created within the application, and by offering visual interactions with the documents contents.

For a system that can organize documents produced with a scanner, analysts should use PaperPort Pro 11 (PC) or DevonThink Pro Office (Mac). For a general purpose active document collector (can be installed on a USB drive) use Mercury document system. Mercury looks and functions much like Outlook, but adds some great search and memo-creating functionality. PDF explorer is great for PDF only collections, since it adds the ability to update PDF Document Description and Document Properties Meta Data. However, as with the recommended Bibliographic Reference Management software, meta-data entry is time and information intensive work. Generally, a policy analyst should not be expected to spend their time filling in data-base fields for their documents.

2.5.6 Content Management

_Evaluated:_ Drupal; Mambo

_Top Ranked:_ Drupal (free, open source: Requires Apache and MySQL)

Content Management tools are primarily enterprise level products, but their availability through the open source community has made them accessible to smaller organizations and individuals. These are web-based products for managing the content of websites separately from their design and technical maintenance. They require some technical expertise and access to install (on a server running Apache and MySQL) but do not require any development or coding to use. They are essentially an online textual
database that is visible to web browsers as a webpage, much like an online forum or a weblog (or ‘blog’).

In general, a policy analyst will not be expected to manage the content of their organizations’ website. Nevertheless, I have included this class of software in this evaluation, primarily because it has been considered the ‘next step’ in the evolution of document management. The principle, explained by Kurtz and Snowden (2003), is the separation of the content of a text from its formatting and display. In a large organization, this allows the technical programmers to take care of making sure the website functions properly, and gives ‘content experts’ direct access to the maintenance and modification of their codified knowledge.

Regarding the individual policy analyst, this class of software has implications for personal knowledge management, since these packages can be used to easily set up a free weblog (Roell, 2004). Efimova (2004) has suggested a range of possible uses of a personal or professional weblog to facilitate personal knowledge management. I have experimented by using my blog [http://mu.tuals.com] as a portable webpage bookmark for when I am working on computers that are not my own, and as place draft my thoughts over the course of my thesis. Given that policy analysis does not rely solely on documentary sources (Bardach, 2000b) but importantly on information from people, a blog could be an important tool for making and recording connections between sources (Efimova, 2004).
2.5.7 Content Analysis, Semantic Network Analysis

_Evaluated:_ Provalis QDAMiner with WordStat; CrawDad Network Text Analysis; AutoMap (unevaluated); Galileo CatPac; Diction; TextStat; Concordance; Copernic Summarizer; TropesZoom

_Top Ranked:_ Provalis QDAMiner with WordStat, CrawDad Network Text Analysis

Content Analysis is primarily a statistical approach to textual analysis (Lowe, 2003). It is a research technique based upon measuring the amount of something in a sampling of some form of communication (Berger, 2004). Busch et al. (2005) identify two varieties of Content Analysis. _Conceptual analysis_ can be thought of as establishing the existence and frequency of concepts – most often represented by words of phrases – in a text. _Relational or semantic analysis_ goes one step further by examining the relationships among concepts in a text (Palmquist, Carley, & Dale, 1997). Their related methodologies are discussed at length by Carley (1993) and Palmquist et al. (1997).

In general, policy analysts do not engage in analysis for the sake of analysis. Purely descriptive statistical analyses of text will not typically be useful to policy analysts. The rationale for including this class of software for evaluation is that it can be used to automatically analyse texts, whereas other analytical tools (like Qualitative Data Analysis) only facilitate manual analyses. However, this is not to say that Content Analysis is necessarily useful to policy analysts.

What could be useful to analysts is the capacity to identify the concepts and their relations in policy arguments in very large collections of texts (Dunn, 1981a), or key elements (Geva-May, 1997b) during the initial phases of policy analysis (Bardach, 2000b; Considine, 2005; Fischer, 1980; Geva-May, 1997b; Howlett & Ramesh, 1995;

The top ranked content analysis packages offer one tool for conceptual analysis, and another for relational analysis. WordStat allows users to create customized coding dictionaries, and provides a range of tools to automatically add related terms. If you have the time to build a ‘dictionary’ of terms that cover your research focus, WordStat can count all the times these terms occur in your collection. It provides semi-automated content analysis using your pre-existing coding dictionary. Alternatively, it can be configured to simply count all occurrences of every term in your collection. If it could use these dictionaries to automatically code the collection in QDAMiner, it would also be the top-ranked QDA tool.

Crawdad reads every text as a network in which each word is a node and computes the network statistics for the most central nodes. It permits batch processing and exports results to an image and HTML file. However, it requires texts in Plain Text format for analysis, and when I tried to run it on my collection of 2500 collected articles and chapters, it crashed consistently. Its developers advised purchasing a supercomputer or trying to run the analysis on any other computer with more memory.

Both recommendations are guarded, given the difficulty and significant time investments involved in the attempt to get these packages to run on my collection of 2,000+ texts. For example, before a collection can be analysed with WordStat, it must be imported into a QDAMiner project file. I found this a very frustrating and time-consuming process. I ultimately had to drop all (N=103) large files (> 10MB) out of my analysis in order to import the files. Once the files were in QDAMiner, WordStat would
occasionally crash without warning midway through an analysis. As my extensive interactions with both WordStat and CrawDad developers could not resolve these problems, the analyses in both software packages will be discussed in a subsequent paper.

In order to use these tools successfully, a policy analyst would need a significant amount of time to create and refine the coding dictionary required in WordStat, and a very powerful computer.

### 2.5.8 Content Summarization

*Evaluated:* Copernic Summarizer; CLRES KMS; Word 2003

*Top Ranked:* Copernic Summarizer

Content Summarization is properly a feature of natural language processing, but its practicality sets it apart. In the domain of NLP, 'true' content summarization involves rewriting a document after classifying it (Farzindar & Lapalme, 2003; Marcu, 2000c; Marcu, 2000b; Marcu, 2000a). While Copernic does not use the most intelligent methods of content summarization (it first identifies key concepts, then picks full sentences that are densely loaded with key concepts) it does a fair job, is simple to use, and exports the results to a handy formatted HTM file. It also provides a weighted list of key concepts drawn from noun phrases. Analysts can use Copernic Summarizer to gain a quick overview of large collections, by reading for gist (O'Halloran, 2003).

### 2.5.9 Natural Language Processing

*Evaluated:* CLRES KMS; VisualText

*Top Ranked:* VisualText
The ostensible goal of Natural Language Processing is to engender in computers a facility with language that will allow humans to address computers as we do other people.

CLRES Knowledge Management System is an elaborate beta program intended for governmental intelligence agencies. It was very difficult to evaluate given its instability. If it were more stable, the range of analytical options it offers would be impressive. However, in its current form, it is unsuitable for either professional or academic use, except perhaps by its designers.

VisualText is a platform for the development of NLP analysis tools. It is used to build textual processing software like automatic Resume Readers. It offers several example programs that developers can use to build upon. For example, it comes with a Part of Speech Tagger that includes a Word Sense Disambiguation module and an automated Named Entity Recognizer. User selection of features or phrases in exemplar texts can train the program by auto-generating and refining rules to assist the software to recognize the general form of the selected case in subsequent texts. For advanced NLP++ programming, some knowledge of the programming language C++ is beneficial. Most policy analysts will have little use for VisualText itself, though they may find certain tools created with VisualText useful.

2.5.10 Software Virtualization

Evaluated: Altiris SVS Admin

34 I wish I had found this software earlier. I might have been able to build the application I have been looking for instead of spending all this time looking. Thankfully, having found this application, I now have a usable framework within which I can go on to build the tool I have been looking for. The future directions of my doctoral research are now feasible (See Section 5.3 Future Research).
Top Ranked: Altiris SVS Admin

Each time my TC1100 failed (See Note 21 in Section 2.4.2 Hardware Description), I had to keep working on other computers while my primary computer was away for repairs. This forced me to see what a wide range of software tools my daily workflow depends on. I have long ago adopted the practice of backing up my data on an external hard drive as a way to protect myself against the possibility of ‘catastrophic data loss.’ But, in order to access my data, I needed software that was not installed on many of my interim computers. I realized that since my work is so tool dependent, I also needed to find a way to protect my data against the catastrophic loss of my system (and the tools installed on it).

Each of the eight or nine times I have received a ‘factory reset’ computer from HP, I had to re-install every piece of software that I regularly use or that was then evaluating. This time consuming process has afforded me the time to think long and hard about what software is actually necessary for my work. The frequent need to start from scratch motivated me to look for tools that could help me to mitigate the time costs associated with reinstalling the necessary tools and ideally, also allow me to take my tools with me, regardless of the computer I have to use.

I found a class of software that allows you to do just that. It is called Software Virtualization. Once you have installed the administrative tool on a PC (free for non-commercial users) you can install any other program into a Virtual Layer that captures all the changes a program installer typically makes on your system. After this, the Virtual

---

35 When the hardware of the physical media on which data is stored (the hard drive) become corrupted or damaged, total or partial loss of recorded data is possible.
Layer can be toggled on or off, and the program appears and disappears from your system. When you turn the layer off and delete the layer, it is as though the program was never installed. It leaves absolutely no traces behind in the file system or, more importantly in the system Registry.

What's more, you can archive and import your Virtual Layers into any system with the admin tool installed. This means you can take your programs with you when you have to use another PC, and remove it when you move to the next one. For the policy analyst who fears becoming 'too technology dependent,' SVS Admin can be used to create a portable collection of core software tools to ensure that you can work with your data even on a backup or borrowed computer.

2.6 Tailored Solutions

2.6.1 Best Fit

The purpose of the ideal computer program at the textual level is to pre-analyse and partially mark-up texts prior to their onscreen display to the analyst. Subsequent to this pre-digestive stage, the software provides a digital workspace for reading, writing, and interacting with texts in a way that has heretofore been impossible.

As the analyst peruses their daily digital textual fare (emails, websites, reports, articles, books, PDFs, Word Documents, Spreadsheets) they will have a number of tools available that they will already be familiar with, from their longstanding tactile experience with paper. Of course, the value-added of this software lies in its integration with the tools and interfaces familiar to the wired generation.
That is to say, traditionally, when an important chapter is photocopied or a seminal article is printed from the internet for the reader to attack with a pen or pencil, highlighters and some sticky-notes; the reader is capturing, processing, and marking-up texts in a fairly non-transferable way. While it arguably adds personal value to the copy, that value cannot easily be transferred among co-workers, or to other associates. Each text marked up in this way becomes, in effect, a highly-personalized data silo. In effect, the original reader becomes the index for that particular copy; without fairly sharp recall, such notes often become meaningless and confusing. Whenever such personalized mark-up defaces the original copy it can make reading more difficult for subsequent readers.

What value can this idealized software bundle add to this traditional analytical reading process? Let me begin to answer this question with another question: What if the traditional annotational tools (pens, highlighters, sticky notes, etc.) had digital proxies that were both technologically feasible and also as easy to use as ‘good old’ pen and paper?

In fact, this technology exists, though the software to make best use of it needs some improvement and integration to make it highly usable. To this end, I have purchased and tested a great range of hardware and software. Perhaps most importantly I have acquired a Tablet PC, which, in addition to the traditional mouse and keyboard, allows its user to interface with the screen directly with a lightweight digital pen. It lets you scrawl notes in handwriting directly on the screen (which is about the size of a pad of legal paper) which can be saved as digital ‘ink’ or converted to typed text. All this is to say that there is currently dependable hardware on the commercial market which uses (and, in some cases, improves upon) the metaphor of pen and paper. However, as yet,
there does not exist textual analysis software that makes explicit use of these intuitive capacities within an integrated suite of textual analysis tools.

To be sure, I have not even been able to find an integrated suite of textual analytic and knowledge management tools suited to the policy analyst. What I have found is that textual analysis is undertaken by a number of disciplines, each with their own priorities for analytical outcomes. Consequently, the tools that I have identified are discipline-specific in their scope and application. In some areas, ranked numerical word counts are considered sufficient to derive statistical characterizations of a given collection of texts. Some areas, like psychological conversational analysis, process transcripts for indications of emotional state. Some, analyse political speeches for indications of tone and other pertinent qualia. Some evaluate focus group or interview transcripts to elucidate themes and topics and to build sociological theory.

In the idealized version, as the computer program proceeds line-by-line through any given text, it will necessarily come into contact with a great many words. However, the literature has established that texts are not simply a random collections of words. Authors carefully select their words and generally place them according to a number of grammatical rules and syntactic conventions. Word order, relative positioning, and declination are each pivotally important structural clues that next-generation textual analysis software must certainly make good use of. It must be able to identify not only the syntactic structure of relations between these words, but also the structure of relations between the concepts they evoke. It must present a dependable means of reducing the sheer number of words, filtering the complexity of their relations, and interpolating between their possible meanings.
Post-structuralist linguists like Werth and Langraker insist that though linguistic expression is intrinsically structured by syntax, sentential structure alone is not sufficient for effective apprehension of intended meaning. In the light of their contribution, it is clear that identifying and recording the grammatical functions of each word in a given sentence (called ‘part-of-speech tagging’ in computational linguistics) is a necessary, though not sufficient step towards identifying textual meaning. However, this and other Natural Language Processing procedures are important steps towards being able to automatically classify the contents of digital documents. I envision a classification schema that draws upon existing NLP procedures to automatically classify documents.

The ideal system for the policy analyst would automatically convert, name, file (move documents to sensible locations), extract meta-data and populate document meta-data fields, classify, group and summarize documents. It should be able to identify structural relationships between documents (as between the chapters of a book, or the articles of a journal, or issues of a journal) and intra-document semantic relations (as of topical groupings).

Ultimately, the system should link individual documents to their Bibliographic Reference meta-data, much as iTunes™ links file meta-data to specific songs. Just in case the analyst cared to develop their own classification system for documents as a whole, (as by topic, publication source or type) or for any individual stretches of text (as with QDA and structured annotation software), this manual functionality should be enabled but not required.
I propose that if Qualitative Data Analysis were integrated with automated Document Management, Content Analysis, Natural Language Processing, Ontology-building, and Summarization software, this would be the ‘killer app’36 for policy analysts.

2.6.2 Practical Recommendations

Having outlined the shortcomings of a range of existing tools, I propose that what is needed to tie together these diverse applications is an information environment (Card, Mackinlay, & Shneiderman, 1999) that offers a range of textual analytical tools (See Section 2.6.1 Best Fit). An important step toward being able to integrate these tools is a common ontology or classificational schema that can be used to automatically classify texts regardless of their disciplinary origins when used in conjunction with POS-tagging. Chapter 3 describes what ontologies are and what they are used for, while Chapter 4 develops and describes the schema I propose.

36 In the subculture of technologist and technology users, the mythology of the killer app-lication mirrors the policy myth of the ‘silver bullet.’ In essence it is the solution that builds demand for the solution. The Jargon files define the killer app as “The application that actually makes a sustaining market for a promising but under-utilized technology.” (Jargon File 4.2.0, 2006)
Note that Knowledge Engineering and System Utilities fall outside of the typical KM tool usage process. The idea is that the creation of an ontology can facilitate other KM and textual analytic procedures; not that the analyst actually spends time creating an ontology. Of course, ontology maintenance remains important.
2.6.3 Future Exemplar of Technologically-enabled Policy Routine

This section provides a future based scenario wherein a Policy Analysts uses both existing tools and strategies (highlighted in bold) and idealized future developments (highlighted in bold italics) to complete a ‘typical’ analytical task.

I started my day early today. Yesterday at the end of the day a request from the Minister landed on my desk. So today I have a policy brief due at 8AM and a briefing session with the CEO of the local Health Authority and Minister of Health at 8:15.

It could have been worse. Even though I am no expert in Health Care, surgery, or robotics, I will deliver the Ministry’s recommendations on co-sponsoring an application to fund the purchase of several systems for robotic surgery. This request required me to become rapidly familiar with very specialized information that was difficult to access, assimilate, and evaluate. I had to collate and synthesize and articles from medical journals, building schematics, organizational financial trends. What’s more, I had to balance the competing perspectives and value systems of surgeons, university and hospital administrators, health authority engineers, Ministry of Health officials, Ministers with various portfolios, and ultimately of citizens, voters, and patients.

And I was able to finish most of it last night and still be home in time for dinner with my family, thanks to the practices I follow, the systems I have in place, and the tools I use.

When first tasked with a new challenge with such a short turn-around time, I hit the road running. But first I follow my own tracks (Bardach, 2000b). I use a desktop search tool (Google Desktop, Copernic Desktop, or Microsoft Live Desktop Search) to
quickly ‘google’ all the documents, emails, and other digital text I have encountered over the years. I also email my top three sources for Health Policy-related issues, asking for their ‘quick and dirty’ lowdown on the topic at hand.

Within seconds, I hit one magazine article I read last year on prostate surgery. I quickly scan the article for people and places, and find the names of several world-class surgeons from Harvard Medical School who are pioneering the technology. I drop their names into a meta-search engine for academic literature (Google Scholar), limiting the results to peer-reviewed articles published in the past 3 years.

At the same time I run a search for individuals of similar or superior social and professional stature to find their qualified peers. I find the names of several more innovators in the field, and locate their recent publications.

Next, I run a keyword search of the recent academic literature and find that my chosen search term produces a range of hits that are not closely on topic. I cannot simply exclude the most common term, ‘laparoscopic techniques’ because some of the most important papers are comparisons between different techniques. So on-the-fly I brainstorm some conceptual groups of policy-relevant issues that might be included in academic publications. These issues broadly include economic, social, and organizational considerations, as well as effectiveness of outcomes for surgery patients. I filter the narrow keyword search using broad conceptual groups.

My new search trims over 2000 articles down to about 100 articles, grouped for each broad issue. I have my system set up to automatically embed all downloads with their original location and the search terms used to find them. I download them all and analyse them for consensus. I do the same for the articles of ‘key players,’ discard the
overlapping articles from the common pool, and compare consensus of the ‘field leaders’ with the rest of ‘the pack.’

The consensus is that while the technique is too new for controlled RCT-type comparisons with other techniques; the recovery time for patients drops to 1 day from 5; initially, the ER time ‘saved’ varies greatly, but the average duration of key operations drops with greater surgeon experience; it is possible to recoup the cost of the units on cost-recovery over about 2-3 years; and finally, the prestige value of having local expertise and training capacity has catapulted previously obscure centres into the world spotlight.

In the course of my academic literature search, I also automatically compiled a list of named entities, like vendors, and automatically collected certain types of published information from them, like marketing materials and white papers. From this smaller collection I gathered the average costs associated with purchasing and maintaining the robots as well as various mocked-up cost-recovery schedules.

A quick review of the Health Authorities’ proposal and supporting documents gave me a rudimentary grasp on the organizational plan for implementing the new technology, and some insight into the attitudes of various parties. I click through the Health Authority’s intranet, examining policy documents and operating room statistics. Every document I browse through is automatically saved locally for offline analysis.

I then reviewed the Ministry of Health intranet, refocused on our organizational priorities, and drafted the briefing note. I was finished in just under an hour. Before I left for the night, I set my analytical software to further synthesize the collection of information I gathered, and then to use my collection as an example of the kind of
information that I want to continue to search for. By the time I woke up at 5:30 this morning, my system had emailed me a further 3 pages of synthesised resources drawn from over 3 million documents in the Ministry of Health, Local Health Authority, and academic Literatures on Health Systems and Policy Research, Operations Research, and Surgical Outcomes.

I feel good about this meeting. I know what I need to know about robotics in our context. This meeting will be a home run!
CHAPTER 3:
WHAT DO YOU MEAN, ‘AN ONTOLOGY’?

3.1 Introduction

This chapter introduces the concept and uses of ontologies. It begins to lay the groundwork necessary to operationalize context in an ontology that can be implemented to automate textual analysis of trans-disciplinary texts (See Chapter 4).

In Section 3.2 I describe the context of domain exclusivity that motivates the need for trans-disciplinary conceptual mapping between discourse from different domains. I then acknowledge the need for verification of internalized conceptions of context against the factual external situations of entities. In Section 3.3 I discuss the extent to which epistemological questions hinge upon ontological answers thereby explaining how efforts to manage knowledge might benefit from the development of a clear, accessible ontology of things and their context. In Section 3.4 I briefly describe several disciplinary approaches to ontology and present several established rationales for developing ontologies. In Section 3.5, I describe the recommended (Noy & McGuinness, 2000) and actual methodology employed in the creation of the schema described in Chapter 4. In Section 3.6 I outline some difficulties encountered when beginning from several traditional starting points to ontological questions.

3.2 Descriptive Realities

Most disciplines, from the scientific to the artistic, aim to describe the world. While their methods and results vary, they universally employ systems of abstraction to
reflect (or reconstitute) various aspects of human experience of reality. While these systems may be widely held within certain cultural or language groups, they are not universal. So not everyone understands them. Many systems are effectively restricted to members of elite groups that require extensive education to gain access (Snowden, 2002). Of course, a degree of domain competence is always required to understand any given system of abstraction. Language is an excellent example of a system of abstraction that requires a baseline competence for effective use. But over time, most disciplines have formalized their own specialized vocabularies and usage of terms; such that extensive domain knowledge is generally presumed of all interlocutors in their discourse domains.

So, regardless of the disciplinary training that readers may have, there will always be some domains to which their training does not provide access, for which readers lack the required domain competence. Domain competence amounts to what Werth and others have called ‘context familiarity.’ My readings across disciplines have affirmed the pivotal role of context in the formulation of meaning (Yanow, 2003; Davidson, Hendrickson, Johnson, Meyers, & Wylie, 1998; Bach, 2000; Crewe & Young, 2002; Gertler, 2003; Hodge & Kress, 1988; King, 1997; Miller, 2002; Pelletier, 2003; Wilson, 2004; Dunn, 2001b; Werth, 1999d; Werth, 1999d). The problem with context is that it is constantly changing, it is specific to each domain, and it is unique to each reader.

While acknowledging the deep educational commitment required to understand disciplinary discourse and the deeply individual way context evolves in the mind of each reader, there is nevertheless a sense in which the totality of existence is not constituted ex nihilo in our minds. None except the most contentious philosophers would contend that everything is in our heads. Rather, the empiricist tradition contends that the extent of our
knowledge of the world depends upon the external existence of the world, our ability to reliably access details about the world through our senses, and our capacity to confirm, invalidate, or augment our conception of things by comparing them with the reality we perceive around us.

Though our conception of context is personal, if we accept the empirical explanation then certain facts of context must be external and verifiable. And though disciplines may use specialized terminology in their description of the world, they are nevertheless describing a world that is accessible to others, who may very well describe the same things differently. Consequently, *trans-disciplinary reading requires the ability to map between different conceptions of reality*, and to know when divergent descriptions are really referring to the same phenomenon.

This poses a great challenge to automated textual analysis. Qualified human readers are able to effectively map between different systems of abstraction, possibly using *mental models* based on their contextual familiarity with the domains in question (Werth, 1999e; Johnson-Laird, 1983; Johnson-Laird, 1998; Knauff & Johnson-Laird, 2002). Any automated classificational system must similarly rely on a *classificational schema* that is in some respects similar to a mental model, but lacks contextual knowledge (Geva-May, 2005). This motivates my attempt, in Chapter 4, to develop a classificational schema, or ontology, for automatically textual analysis that can categorize things and their context.
3.3 Epistemological Implications

Even while acknowledging the subjectivity of personal knowledge and perspectives, it is often very difficult to see how our assumptions about reality impact our ideas about the world. It is not easy to identify the existential assumptions that underlie our way of thinking. It is even harder to describe their effects on our ideas about knowledge. Given the difficulty of this exercise for the average individual, any systematic attempt to delineate the existential assumptions upon which a disciplinary or traditional way of thinking are founded is sure to be nearly impossible. However, this is precisely what we need to do in order to create an ontology for things and their context.

Regardless of the difficulty involved, our thinking about what exists and what can exist significantly determines our thinking about what knowledge is and what can be known (Kurtz et al., 2003). If we can clear up some basic questions about the nature of existence, or at least surface our implicit assumptions about existence, we stand to clarify our understanding of knowledge, what it is, and how we can use it productively (Kurtz et al., 2003). Understanding our conception of existence can provide insight into how we gain our knowledge of things, and ultimately what counts as ‘knowable’ (Yanow, 2003). Consequently, when we organize our assumptions and ideas about what exists into an ontology, the resulting ontology must be capable of accommodating all the things that are known to exist in the scope of the ontologies’ domain.

There are key implications here that draw from earlier discussions of ‘theories’ or metaphors used to describe knowledge. For the purposes of this discussion, knowledge the residue of human experience, processed and evaluated sense datums that are drawn into relation with prior experiential knowledge, in concert with accepted or learned
systems of abstraction and their contents that are not necessarily personally experienced but that are nevertheless accepted on the basis of evolutionary or explanatory value. For example, language is not a sense datum; but it is an abstract system that is experienced, and consequently learned. The contents of discourse are not experiential datums, per se, but they may recount those of others (or even other language-encoded systems of abstraction, like theories or mathematics). The experiences or theories of others can be accepted, as with sensory data, provided they fit with previously held ‘knowledge.’ When they do not; they are usually qualified or discarded; though their acceptance may cause an individual to re-evaluate and even reject their previously held knowledge.

3.4 Ontological Flavours

To a point, language systematizes and aligns existential assumptions enough to facilitate communication. When we speak, we do not think of itemizing and qualifying all of our implicit assumptions. We generally use what we presume to be definitive gestures towards things or ideas that we do not think we need to justify. For example, grammar provides us with a limited set of categories (Nouns, Verbs, Adjectives, Adverbs, etc.) and syntax rules for their proper combination. Our use of language amounts to tacit agreement to a de-facto, basic, common ‘ontology’ of ‘persons, places, things, and ideas’, ‘actions’, and ‘descriptions’ thereof. By using language, we attest the existence not only of the things to which our words refer, but also of the existence of these linguistic categories; and by their usefulness, of their necessity.

In philosophy, ‘ontology’ is the systematic study of existence. This is most often achieved by identifying its basic divisions and its resulting constituents. Since most ontologies are a categorization of existence, it is important to draw a distinction between
the categories and those things that ostensibly belong in them. In philosophy, this distinction is reinforced by using the term ‘universals’ for those things (categories, classes, kinds, types) that can have instances, and ‘particulars’ for those that cannot. Particulars are the real-world instances that universals serve to classify (Oltramari, Gangemi, Guarino, & Masolo, 2002).

In computing science, an ontology defines a common vocabulary for researchers who need to share information in a domain. It includes machine-interpretable definitions of basic concepts in the domain and relations among them. Noy and McGuiness (2000) report that many disciplines now develop standardized ontologies that domain experts can use to share and annotate information in their fields. They go on to suggest several reasons for developing an ontology, each of which have motivated my own ontology-building effort:

- To share common understanding of the structure of information among people (or software agents)
- To enable reuse of domain knowledge
- To make domain assumptions explicit
- To separate domain knowledge from operational knowledge
- To analyze domain knowledge (Noy et al., 2000)

The ontologies developed by information architects for the ‘semantic web’ are intended to render ‘common-sense, everyday relations’ into a machine-readable format by representing them with formal notation developed in logic and mathematics (Guarino

---

37 According to Masalo et al.s. (2002) the idea of the Semantic Web hinges on the possibility of using “shared vocabularies for describing resource content and capabilities, whose semantics is described in a (reasonably) unambiguous and machine-processable form.”
et al., 2004). However, the purpose of the schema that I develop is not to describe semantic relationships using formal notation. Instead, I developed a framework to describe ‘things in context’ using natural language in a multi-layered matrix using the common spreadsheet application MS Excel™. While my approach does not begin to formally describe the relations between categories in mathematical or logical notation, this is a direction for future development.

An excellent example of a system that uses formal, first order logic (calculus) in concert with a knowledge base structured by a ‘top level’ ontology is called Cyc. This project has been in progress since the mid nineteen-eighties, and is an attempt to teach computers ‘common sense.’ Their approach is to translate ‘obvious’ facts about the world into a highly structured ‘formal’ language that is understandable to computers. They have spent millions of dollars and over 20 years training their system this way. The chief problem with this system and approach is that it requires a high degree of competence in calculus in order to convert common facts into machine-readable code, so training the system to ‘extend’ its knowledge base is very time consuming. Policy analysts simply do not have time to ‘feed’ this kind of system in this way that is currently necessary to train this type of General Artificial Intelligence.

Other examples of applied ontologies that could be useful to Policy Analysts include Pub Med’s Medical Subject Heading (MeSH) Terms. MeSH terms are a controlled vocabulary used by a select group of librarians to describe the published medical literature. In this senses, they are a domain-specific application of what librarians

---

38 For a long while, I also aspired to unpack all of the hidden assumptions and existential dependencies of any given concept in my framework. I have since retreated from the stringency of this approach. I now admit the possibility that lexical definitions do not necessarily contain logical predicates of terms, but rather existential co-requisites. (Rockwell, 2005c)
already do; identify which ‘official’ keywords should be used to describe the contents of
a published document in the library holdings. Both systems can be useful to Analysts
looking for texts. However, these systems also have weaknesses. These drawbacks also
parallel current weaknesses with internet search technologies; namely, that users are
constrained to matching their search terms to either the Subject Headings, or to words as
they actually appear in online resources.

It is currently not possible to quickly and simply search using conceptual groups;
unless you are familiar with the domain and its terminology. Even then, it is often
difficult for users to bridge the gap between how they would describe what they are
looking for and how various disciplines may conceive of those resources. This is one of
the problems the ontology described in Chapter 4, once implemented, may help to solve.

3.5 Ontology-Building Methodology

Noy and McGuinness (2003) describe the process of building an ontology as a
matter of identifying the classes of things that exist within a domain of interest,
identifying the properties that distinguish these things, and identifying an exhaustive set
of instances to typify these things and properties. This is the generic process that I have
followed. However, while I have chosen my starting point carefully, I have identified
both classes and their instances with explicit reference to the lexical definitions of terms
(See Sections 4.4 through 4.6 for detailed explanations of the derivation of the 12 top-
level categories in this schema. See Section 4.6.6 Requisites for further clarification on
the method used to populate the schema with instances.)
3.6 Troublesome Beginnings

All questions of existence must start somewhere. Standard folk accounts begin in a narrative fashion by situating things in a historic context. This highlights the valuable intuition that accounts of what exists today depend on what came before. However, on reflection this approach presents several problems. First, history constitutes a limited account from a fixed perspective. Many more perspectives with many more details are not a part of recorded history. Also, for every narrative there always remains an era before observations were systematically recorded.39

Other more formal philosophical accounts also begin with time. Their premise is based on a fundamental distinction between the behaviour of ‘enduring’ and ‘perduring’ entities in time.

Endurants are wholly present (i.e., all their proper parts are present) at any time they are present. Perdurants, on the other hand, just extend in time by accumulating different temporal parts, so that, at any time they are present, they are only partially present, in the sense that some of their proper temporal parts (e.g., their previous or future phases) may be not present. E.g., the piece of paper you are reading now is wholly present, while some temporal parts of your reading are not present any more (Masolo et al., 2002).

Philosophers say that endurants are entities that are in time but lack temporal parts (this is to say, all their parts flow with them in time). Perdurants, on the other hand, are entities that happen in time, and can have temporal parts (all their parts are fixed in time). So both can be characterized by their ability to exhibit change in time. Endurants can

---

39 The rationalist impulse assures us that simple conjecture about prehistory cannot stand up to the presumed rigour of recorded and validated observations. The resulting dilemma is that, lacking that original observation and its subsequent validation, we can never satisfactorily explain the origin of ‘what is.’ And, even where recorded observations have endured in a manner that remains accessible to contemporary readers, there remains a question as to the readers’ ability to correctly interpret the intended meaning. The problem of temporal and cultural distance for interpretation of texts is address in the discipline of Literary Hermeneutics (Robinson, 1995).
“genuinely” change in time, in the sense that the very same endurant as a whole can have incompatible properties at different times; perdurants cannot, since their parts do not maintain their identity in time (Oltramari et al., 2002).

More traditional ontologies begin with the assumption of another particularly classical binary. All things must fall into one of two initial categories. Classically, they are either material or immaterial; substantial or insubstantial; concrete or non-concrete. The first problem with this approach is that it builds in a materialist bias. Binary logic requires that where the first of a dichotomous pair is described as being or having something, its counterpart must be or have the opposite. Matter is observable. Substance is manipulable. The concrete is tangible. So (if and only if [iff] synonymy relations hold between the given terms’ antonyms) immaterial things are necessarily intangible, non-manipulable, and unobservable.

This approach is problematic in that it leaves little room for generating or validating descriptions of non-material phenomenon, except where perfect material oppositions exist. If abstracts are indeed unobservable, intangible, and non-manipulable, how can we describe them? How can we grapple with abstracts that have no perfect, opposite, material manifestation?40

These questions highlight how the initial existential assumptions of an approach can circumscribe the questions that can be asked and ultimately the answers that they can reach. If we begin from here, the terms of this debate would neatly circumvent discussion of much of what potentially exists (the ‘abstract’). This starting point could

---

40 Another problem regards how frequently binary oppositions map to complementary concepts that are not strictly opposites. This question regards the proper binary opposition for any given category. Should the binary opposition be a lexical antonym or a logical negation? (See Appendix D: Determining Binary Opposites)
conceivably also limit our capacity to address the characteristics attributed to things as 'properties,' insofar as properties are an abstract description of a feature that may not even necessarily be physical.

While the distinctions between opposite pairs of properties like concrete and abstract or enduring and perduring do play a very important role in this schema (See Section 4.6.2 Ranges), they are not the starting point for the ontology. My approach is based on a pragmatic phenomenology of human experience (See Section 4.3 Experiential Requirements). I use this approach to identify a categorization schema for both the primary kinds of things that exist and the different kinds of situations in which these phenomena occur.
CHAPTER 4: WHAT DOES THE SCHEMA OFFER?

4.1 Introduction

This chapter describes a schema that intends to accommodate the divergent perspectives that underlie descriptions of the world found in disciplinary texts. With this schema I hope to address some of criticisms I have drawn against Knowledge Management theory and textual analysis products evaluated in this thesis. First, I intend to be explicit about the nature of both existence and knowledge. While I do not presume to have solved any age-old epistemological or existential dilemmas, I believe I have found a viable way forward.

In Section 4.2, I reflect upon the need that ontological questions evince of a particular kind of mind that is capable of certain mental processes. In Section 4.3, I begin to try to operationalize ‘things and their context’ by thinking about human experience of the self and the world. In Section 4.4, I reframe my approach as explanatory and descriptive, briefly outline my way forward, and begin to clarify my initial assumptions. In Section 4.5, I introduce the Categories of Things and explain their derivation from my initial assumptions. In Section 4.6, I propose the Elements of Context, explain their derivation from human experiential phenomena, and describe each in detail.
4.2 Abstract Processes

I intend to build a classificational system that explicitly acknowledges its existential assumptions, dependencies, and prerequisites. Given the philosophical roots of any ontological endeavour, its first challenge should not come as a complete surprise: Questions of existence are intrinsically a mental exercise. How, then, can we begin along this line of inquiry when the apparatus that enables this activity depends upon the existence of so many other things? Where can we start that does not presume the existence of things? Given that our existential presumptions seem to be unavoidable (Rockwell, 2005c), I argue that the responsible theorist must attempt to reduce and explain her schema’s existential dependencies.

Classic ontologies derive from an epistemologically assumed ‘object’ (something that is presumed to be known) that somehow manifests its most salient features into categories called ‘properties.’ They just assume that categories exist. They do not question how they have come to exist, or what they might be. They do not suggest or hint that categorization is a mental process that mobilizes perceptual, observational, comparative, and rational processes, or that categories only exist insofar as we have created them using the kinds of minds that we ostensibly have (Churchland, 2002; Kolb & Whishaw, 2001; Marcus, 2004).

This is not to say that the existence of things depends upon us having our kinds of mind, just that the way we identify things according to their manifested properties does. Categories are, in fact, an abstraction. They are the result of the processes of observation, identification, recollection, recognition that all occur in the material, biological substrate of our bodies and brains.
Historically, what precisely transpired in our brains as we process perceptual data was not easy to observe (Masolo et al., 2002; Rockwell, 2005a). However, advances in medical imaging and neurophysiology have rendered many of these perdurant electrochemical phenomenon observable (Kolb et al., 2001). Of course, this thesis is not principally concerned with the details of neural functions and pathways. Nevertheless, these details offer a starting point to the important existential and philosophical questions that underlie this work.

The human body is calibrated to receive, collate, and organize information about the world through its nervous system. The sense organs are dedicated channels to the outside world that continuously feed our brains with details about our surroundings. Some of these channels are divided hemispherically by our physiology, so our brains must interpolate two slightly different streams of information. Amazingly enough, this does not pose a significant challenge to healthy bodies. Our brains appear to be remarkably well-adapted to accommodating divergent perspectives (Kolb et al., 2001). In some cases the different streams are unified, as with our visual field (Milner & Goodale, 1995); and in others we perceive their difference, as with our audible field (Allefeld, 2004; Mori & Kai, 2002; Mori et al., 2002; Mori, 2002).

4.3 Experiential Requirements

This schema intends to frame existential dependencies explicitly in terms of human experience. Accordingly, this section describes the existential dependencies of

---

41 According to the Behaviouralist tradition in Cognitive Science, much as our senses are channels to the outside world, our sensory-motor behaviours are indications of the internal world of instinctive or intentional thought (Kolb et al., 2001). This fact is important to certain slots in this schema (discussed further in Sections 4.5 and 4.6).
things starting from the thinking, categorizing human. As the foregoing section has established, the experiences of healthy, socialized human beings are initially modulated by their physiological and neurological systems. In other words, human experience requires a body (particularly the kind of body with a central nervous system equipped to handle external sensory stimuli and a brain adapted to manage it all). Bodies are unique configurations of matter that have the property of being alive.

So human experience requires a body, and bodies requires matter. Matter is a relatively stable state of energy that occupies space and can be perceived by the senses. So human experience requires a body and matter; which requires space.

Space is the infinite extension of the three-dimensional region in which all matter exists. Space depends on the orderly intersection of individual linear dimensions to form planes and spaces. So, experience requires a body, matter, and space; which requires planes, which require lines.

But human experience entails things like sensation, perception, and living, which are all actions and require the possession of properties that change. Action and change require time. So human experience requires a body, matter, space (planes, lines) and time.

Human experience requires sensory channels to deliver perceptual data to the brain, which is capable of a range of mental processes at various levels. So experience requires a mind; which requires a body, matter, space, planes, lines, and time (Rockwell, 2005c).
But as mentioned previously, some sensory channels are split, offering divergent perspectives on any given scene. Moreover, the bodies in which our sensory channels are embedded are ambulatory and capable of a wide range of motions through space and time. Our observations of situations change according to our spatio-temporal and mental positions. Accordingly, perspectives are the sum of our perceptual and reflective trajectory through space, time, and mind. So human experience requires perspectives; which require a mind, body, matter, space, planes, lines, and time.

The human experience is not primarily one of solitude. Patterns of production and reproduction usually result in continuous and varied interaction among individual humans. Society is the totality of relations between humans and their enduring constructs, like institutions and cultures. Humans experience the world from a truly unique perspective, informed not only by our own trajectory through time and space and by our preferred way of thinking about things (or perspectives), but also by the views and accepted accounts of those individuals and institutions we have learned to trust. So human experience requires societies, which require perspectives, minds, bodies, matter, space, planes, lines, and times.

According to some theories of evolution, human interactions were initially predicated as a mutually beneficial response to survival pressures (Gardner, 1999; Gardner, 2000). Possibly in order to work effectively together, humans adapted and systematized the communication of thoughts and feelings through a system of arbitrary signals, such as voice sounds, gestures, or written symbols (Kolb et al., 2001). As such, modern human experience requires language; which requires society, perspective, mind, body, matter, space, planes, lines, and times.
For many, the human experience has extended the evolutionary quest for survival into a quest for ‘self actualization’ by finding meaning and value in life (Maslow, 1970). Appreciation of the subtle connections between things, symbols and meanings marks what some have called the spiritual phenomena of understanding (Emmons, 2000; Gardner, 2000; Kwilecki, 2000; Mayer, 2000; Vaughan, 2002; Zohar & Marshall, 2000). Accordingly, human experience requires the spiritual; which requires language, society, perspective, mind, body, matter, space, planes, lines and times.

Ultimately, the human experience of all these phenomena depends on the capacity to identify and recognize individual things and their properties. But the concepts of individuality or plurality require a concept of Number. However, numbers belong to a class of things that are purely abstract, in that they have no physical instances. Our ability to recognize them is based entirely on comprehension of abstract concepts. So pure abstractions require not only minds; but also understandings. Yet since the concept of numbers is paramount to identifying the boundaries and identities of non-abstract, material things as well as phenomenological perdurant things, human experience requires the mathematical, which requires the spiritual, linguistic, social, perspectival, mental, corporeal, material, temporal, spatial, planar, linear; all of which require the mathematical.

Accordingly, in the proposed schema, mathematics, as the domain of pure abstraction, is the keystone in an existential loop that begins and ends not with the mind as such but with its ultimate product: Abstraction. Our capacity to recognize differences

---

42 Over time, as our capacities for comprehension deepen (or languish), our experiences of our engagement in the world create (or diminish) capacity to engage productively and effectively in the world. In this way, our view of the world reflects and actively creates the meanings and senses we get of the world that we experience. Our memories of our own experiences and our experiences of learning things sec
between situations in which phenomena occur; or to track the existential dependencies of successive *Categories of Things* is premised upon our ability to recognize and identify individuals and to group them based upon their possession or lack of properties. So, human experience relies upon the phenomena of mental processes, corporeal bodies, material substances, spatial locations, temporal occurrences, perspectival constructions, social cultures, linguistic conventions, spiritual capacities, and mathematical abstractions. Consequently, human knowledge rests upon human experience and the associated capacity to accept abstractions generated by others. See Section 3.3 *Epistemological Implications* and Section 1.3 *Ideal Formulation* for further discussion of knowledge.

4.4 Phenomenological Schema

The preceding sections have established that actively questioning the nature of existence is no easy task. In my experience, the process has lead me to questioning the grounds of my own assumptions. To date, my greatest difficulty has been justifying my assumptions about the existence of basic things, like entities, properties, or categories.  

These difficulties compound when attempting to build a schematic representation of an ontological framework. So the schema described in this paper starts from a different premise. I sidestep my largest hurdle by identifying the experience of the categorizing human as the starting point for my ontology. Since the ontological question (What exists?) has remained unresolved since early Antiquity (Guarino et al., 2004), this paper

\[43\] Say, for example, you begin by asserting the existence of 'obvious' material things and follow with some sort of corollary about non-material things. Implicitly you must assume the existence of categories and of instances; and that 'properties' are an instance of a category; and further that 'material' is an instance of 'property.' In any case, you start you end up invoking things which you have not (and apparently cannot) prove to exist before making use of them in your existential explanations.
will not attempt to justify or prove existence of things, but rather to explain them in terms of their relations.

By explain, I mean to systematically describe a framework that can accommodate all facets of human experience using categories derived jointly from heuristics for human language understanding and from a phenomenology of human experience. My purpose in this section is to describe how I have derived the schema. I begin by acknowledging my significant initial assumptions. I then go on to explore some of the syntactic and semantic connections of conceptual word groups and develop a psycho-linguistic structural framework for categorizing their contents.  

**Stated Assumptions**

1. Things exist.

2. Relations situate all things.
   a. Parts belong to Wholes.
   b. Features belong to Identities.
   c. Individuals belong to Groups.
   d. Instances belong to Categories.
   e. Things belong to Contexts.

3. All things that exist in reality are Entities.

44 Having noted my assumptions, I may subsequently elect to return to attempt their justification. However, since the purpose here is exploratory and theoretical, formal justification of the grounds for these assumptions is not the measure of success for this exercise.
In its classic interpretation, claim (1) asserts two things: there are things, and they have properties (like existence). Note how this interpretation uses a vague locative term (‘there’) to situate things and how it somehow collapses a present-perfect verb (‘exist’) into the possession of a property (‘having existence’). Also note that terms are discussed in their plural forms.

These details are each interesting in their own right. The first is evidence of an impulse to situate or ground even generic ‘things’ with a prototypical notion of location (Cook, 2004). This highlights the centrality of Space in our innate sense-making rubric and partially motivates claim (2). Plural terms highlight the importance of Number, suggesting that existence is not conceived of as monolithic or homogenous; but rather as comprised of many things. The implication is that numbers serve several important roles, as by counting individuals and constituting groups.

The linguistic transmutation of a verb form into a property involves a few more steps. It first involves stepping back from the original claim. Given that claim (1) is
offered in a particular linguistic form, it can be rewritten to express roughly the same idea (There are things; Things have existence). The problem is that these reformulated sentences make more explicit references to a greater number of uninstantiated ‘things.’ (On principle, I want to begin the ontology by assuming as few things as possible.) As it stands, the original formulation of claim (1) already implies the existence of Existence, Numbers, and Actions.

So while I elect to keep the original claim, note how each phrasing conceives of existence differently (Spatially, Actionally, and Possessively) and how these differences follow from their part of speech. In the first it is an Action; the second it is a State of being; while in the third it is a Property. On the surface, it would appear to be simply a matter of linguistic preference as to which to ‘kind of thing’ you conceive existence to be. My choice is motivated by principled concision.

4.5 Existential Categories

These linguistic permutations offer us a wider vantage point from which it is conceivable that ‘existence’ could simultaneously be an entity; a property; a state; a power; an action; and the outcome of that action. That is to say; claim (1) implicitly asserts that existence is a thing, that things have the property of existence, that they are currently in a state of existence whenever it explicitly states that things are presently doing some action called ‘existing.’ Common-sense confirms that the completion of an action amounts to the production of an outcome (which, in this case, is simply the reinforcement of a state). To apply some post hoc reasoning; knowing that an action is in progress or has been completed necessarily implies that the capacity to act must have
existed prior to its enactment. So things that exist necessarily have the power to exist and inevitably produce the outcome of existing.

Existence is just one property that things can have, state that things can be in, power that things can actualize, action that things can do, or outcome that things can cause. As such, existence is an instance that can belong to each of these categories. But how can one ‘thing’ can be a property, state, power, action and outcome?

This insight is clarified by defining a ‘concept’ as a loose group of terms that share a core lexical reference. We have a ‘concept’ for each ‘thing’ that exists in the world. The individual words that belong to these ‘conceptual word groups’ each render a unique sense of the core reference. So, for every concept that can be expressed in language, a range of related words exist to express different aspects, or facets, of the given idea.

For example, the terms ‘concept’ and ‘conceptualization’ are nouns; ‘conceptual’ and ‘conceptualizable’ are adjectives; ‘conceptualize’ is a verb; and ‘conceptually’ is an adverb. Each refers to the core lexical reference, but their morphological form and grammatical part of speech serve to structure and limit which facet each refers to. They all refer to different facets of the same core reference; ‘The Concept.’

These words are connected to each other by merit of their relation to that central reference (and, as we will see, by their membership in a set of related categories). Their plurality is owed to the fact that a single form of the word would not be enough to express the variety of our human experiences with the core idea. The multiple facets of the concept allow us to describe the ‘thing’ variously as an Entity, Property, State, Power, Action, or Outcome.
Recall (from Section 3.4 *Ontological Flavours*) how philosophy assures us that the *Categories of Things* are quite different from the *Instances of Things* that they contain. We can reinforce this distinction by using terminological variations of the category names when we want to refer to the instances that belong in those categories. As such, a *Prototype* is an instance of an Entity; a *Quality* is an instance of a Property; a *Condition* is an instance of a State; a *Capacity* is an instance of a Power; an *Act* is an instance of an Action; and a *Result* is an instance of an Outcome. The assumptions describing the relations between *Categories of Things* are outlined below.

**Stated Assumptions**

4 Entities have Properties.

5 Properties describe Entities and instantiate States.

6 States require Properties and describe Powers.

7 Powers require States and enable Actions.

8 Actions require Powers and produce Outcomes.

9 Outcomes require Actions.

**Unpacked Implications**

 Assumes Possession.
 Assumes Instantiation.
 Assumes Description.
 Assumes Requirement.
 Assumes Enablement.
 Assumes Production.
These *Categories of Things* were not randomly selected, nor were they achieved by simply renaming familiar grammatical categories. In fact, they were identified through an arduous process of manually mapping the lexical definitions of nearly 3000 terms (For a description of this process, See Section 4.6.6 Requisites). They represent synthesis of similarities and differences drawn from numerous senses listed in publicly available lexical resources (including the various dictionaries and WordNet 2.0 available on dictionary.com).

These ‘top-level’\(^{45}\) *Categories of Things* reflect a continuous view of existence that draws its categories into causal and dependent relationships with each other. This is important, since it allows for inferences to be made about the ‘thing’ once a term is identified as belonging to one of these categories. For example take the sentence: “The donuts are fattening.” Knowing that fattening is an adjective helps to identify the term as a property of these doughnuts. Given the relationships between categories we can infer that by possessing this property, these donuts are in a state of being that involves the potential to accomplish an action that could create the outcome of actively fattening something else (making X fat). Knowing that donuts are usually eaten provides a conditional trigger that can clarify and limit the impact of their ‘fattening’ potential to their eaters (they do not fatten themselves, nor do they fatten everyone).

However, as a twist on the traditional conception, this continuum is conceived of as circular rather than linear. This insight is premised on the fact that though we arbitrarily ‘begin’ with the existence of things and proceed onward to identify their

---

\(^{45}\) These *Categories of Things* are considered ‘top-level’ categories, because (barring the original bifurcation between Things and Situations) they are the most basic distinctions drawn between things in this schema.
properties, states, powers, actions and outcomes; in fact all entities are outcomes. The implication is that the existence of things is an outcome of some action, which was presaged by some power, which was possible in some state, which resulted from possession of some property, which belonged to some thing. We can use the relations between categories to make reasonable inferences (in either circilinear direction, or even across, to different points on the circle, See Figure 7, Phenomenon of human Experience,) from whatever starting point we are given. This circular conception highlights the bounded nature of human reflection upon existence; wherein some innate scope is embedded in human knowledge of things. There is a point beyond which knowledge of the origins and even the existence of things is necessarily fuzzy, uncertain, and partly unknowable.

Figure 7. Phenomena of Human Experience

The logical and progressive relations between these categories form the basis of the first half of my ontology and serve to classify the Phenomenon of human experience.
4.6 Contextual Elements

However, my intention is to create an ontology not only for entities but also for their context. The principle motivating this endeavour is that all things are situated. Recall for a moment our first two stated assumptions (and their corollaries). Claim (1), the *Principle of Existence*, holds that Things exist.\(^46\) Claim (2) asserts that Relations situate all Things. Implicit in claim (2) are the assumptions that Relations belong to the generic class of ‘things’ (A Relation is a Thing); that Relations have the property of existence; that Relations serve to situate things. Claim (1) relies upon the concept of Existence, while claim (2) depends upon the concept of Situation.

Claim (2e) maintains that Things belong to Context. Implicit in claim (2e) are the assumptions that Context is a Thing that has the property of existence and that Context is an assemblage of Things (much as a Whole is a composite of Parts; an Identity is a compendium of Properties; a Group is an aggregation of Individuals; or a Category is a collection of Instances). Claim (2e) depends upon the concepts of Belonging and Number. As mentioned previously, Number (implied by the use of plural terms) is necessary for a concept of Individuality, Plurality, or Groups. (It further implies the existence of Properties, Differences, Similarities, and Identities; for only if a things’ Identity is known can a pair or Group be formed; and only when its Properties are recognized in terms of Differences and Similarities can its Identity be determined.)

Of the many different Relations that exist to describe the variety of possible connections between things, the concept of Belonging describes one such relation from the perspective of the individual members of a group. (The concept of Possession

\(^46\) The concept of Existence is described in the foregoing section. Situation is a manner of positioning with reference to surroundings. In its prototypical usage, it is about relationships of things in space.
describes the same relation from a perspective the owner of the group.) So while Context may be conceived of as a ‘super set’ of Things and their Relations; it is also, somewhat paradoxically, an instance of both the ‘Belonging’ and ‘Possession’ relations.

The *Principle of Situation* uses the term ‘Context’ to describe the situatedness of things. In this view, context is both an environment or milieu and the sum of the relations between entities in any given situation. Though this may seem to present an impossibly grand scope, having just identified 6 categories to which ‘all things that exist’ ostensibly belong, we can tame the perceived expansiveness somewhat by considering how the *Principle of Situation* manifests for each *Category of Things*. What things situate Entities? Properties? States? Powers? Actions? Outcomes?

We have already begun to answer this question in the foregoing section by establishing the basic relations that hold between top-level *Categories of Things*. Now we must focus on identifying those relations that might hold between the instances belonging to these categories. Just imagine that we need to refine the massive and ever growing collection of Entities we have successfully used the first half of our schema to identify.47 How might we usefully differentiate between the different kinds of Entities that exist? To start, we can derive and build a classificational schema from our human experience of phenomena (As developed in Section 4.3 *Experiential Requirements*) that will serve to situate Entities.

We might ask, with the intention of classifying the different kinds of situations in which humans experience the phenomena of things, what *Elements of Context* must

---

47 Recall that the Entity slot in this schema is reserved for the noun form of the term used to describe a concept (Situation). No matter what form of the term we are given, we can usually identify (or fabricate) its related terms (Situated, Situating, Situtable, Situational are Properties; Situatedness is a State; Situating is a Power; Situate is an Action; Situations and Situations are Outcomes).
necessarily exist in order for humans to experience things the way that they do. This initial stratifying concept requires a name.

Recalling the importance of Space in situating the original Principle of Existence, I was initially inclined to look for the Domains (or some other variation on the spatial concept) of Context. However, on reflection I realized that, as a composite of three linear Dimensions in a relationship of mutual perpendicular intersection, Space can be reduced to a more basic element: Dimensions. Acknowledging the importance of space (without assuming that all contexts are necessarily spaces) to the concept of Situation, I identify ‘Dimensions’ as the first of six Elements of Context.

Figure 8. Elements of Context situate Categories of Things

Stated Assumptions

10 Dimensions situate Entities.

11 Ranges situate Properties.
Modalities situate States.

Extents situate Powers.

Orders situate Actions.

Requisites situate Outcomes.

**Unpacked Implications**

Assumes Situation.

Previously, I introduced Snowden’s Cynefin model of Context. However, Cynefin only includes two dimensions of context: Culture and Abstraction. While these dimensions may be adequate for explorations of the formulation of knowledge in an organizational context, they are not sufficiently diverse to accommodate a fuller notion of context as experienced by humans in everyday life; or by policy analysts in their daily work.

Originally, I attempted to build this second part of the schema by treating every word in a terms’ definition as a pre-requisite for the existence of the defined concept. So, since the definition of Dimension implicated Measures, Extents, and Magnitudes, then these three things ‘must’ be shown to exist before I could presume that Dimensions exist. The obvious problem with this approach is that the definitions of most terms use many words to describe many senses of a given defined term. So, while I had intended to select my ontological entry-points very carefully, using this approach I was obliged to ‘pin down’ the existence of an exponentially growing list of concepts before building my schema. Yet I quickly found that the only way to I could begin accommodate all these ‘implicated’ concepts was with a schema.
I eventually realized that my problem was that I had unconsciously decided to take definitional terms as existential prerequisites. This assumption presumed an existential order that reflects common dictionary usage, whereby the user seeks to define a concept in terms with which she is already familiar. But this process reflects epistemic order and not necessarily existential order. Nevertheless, by the time that I reached this conclusion I had already identified and mapped instances for half of the Elements of Context (e.g. Extents, Orders, and Requisites) as the existential requisites of Dimensions.

Consequently, these three Elements of Context initially served to gather and organize the existential pre-requisites of Dimensions. However, on closer inspection, all six Elements of Context actually serve quite well at situating not only specific categories of things but also any given category of things. That is to say, all things (Entities, Properties, States, Powers, Actions, Outcomes) prototypically belong to dimensions and have specific modalities, extents, orders and requisites. But Modalities (especially as existential modes of being) intuitively go hand in hand with States; Extents help to demark the otherwise nebulous concept of Powers; Orders index Actions; and Requisites are necessary precedents to any Outcomes. So all Elements of Context have both a prototypical and a universal scope of application.

4.6.1 Dimensions

A Dimension is a measure of extent or magnitude. According to my initial assumptions, in order for Dimension to exist; Measures, Extents, and Magnitudes must exist. Accordingly, a measure is defined as a conversion of a property of a physical entity or an event in nature into some vicarious symbolic system (like math or language). An extent is the utmost point or limit of capacity. A magnitude is a relative property of size
or extent that can be described by a number. The implications of these definitions suggest that a Dimension must be able to describe and compare the full scope of things, properties and relations that exist in its purview.

It may not be intuitively apparent at first glance how Dimensions can serve to stratify human experience. As defined by the dictionary, Dimensions are meant to describe or compare properties of things in a given extent according to an established standard. Dimensions are measures of things. For their part, Measures are intended to comparatively enumerate the properties of things that properly belong along or exist within a given extent. So Dimensions allow us to grapple with things by systematically converting their properties into values in recognizable symbolic systems and by indexing them to the situation in which they exist.

At first glance, very few things are immediately recognizable as Dimensions. Time and Space are the two most recognizable Dimensions that effectively serve as primary differentiations between things. But not all things are either temporal or spatial; so more Dimensions are needed. For example, Numbers are a different kind of thing than Moments or Locations. While Moments are temporal things and Locations are spatial things; Numbers are a mathematical kind of thing. So while the intuition to extend the concept of Dimensions to the realm of Numbers and mathematics is easy to follow; we are still left to determine the relationship of this mathematical Dimension with those of time and space. Since the concept of individuals and groups is pivotal to classifying things, and given that these concepts are a logical outgrowth of Numbers, I locate the mathematical Dimension of Numbers just prior to the dimensions of Space and Time.
Then there are the differences between physical, corporeal, mental, perspectival, social, linguistic, and spiritual kinds of things (See Section 4.3 Experiential Requirements.) In this schema I propose a separate Dimension for each. As suggested in Section 4.4 Phenomenological Schema, the things in each successive Dimension depend on those in the preceding ones. The Material Dimension depends on space and time. The Corporeal Dimension animates the Material; and the Mental illuminates the Corporeal. In the Perspectival Dimension, individuals absorb and reflect their unique perceptions of the world. In the Social Dimension individuals negotiate the dynamics of group life; the Linguistic Dimension structures the possibility of highly effective communication; while in the Spiritual Dimension individuals seek and find esoteric and every-day meaning in the experiences of life.

Recall (from Sections 4.3 and 4.6) that Space is in fact a composite of three dimensions in a particularly dependable relation; the Linear, the Planar and the Spatial.  

48 Colors highlighting each Instance express prototypical Dimension. See Fig.11 Dimensions of Existence.
This highlights two things. At its roots, the concept of a Dimension fundamentally depends on the concept of a Line. *So Dimensions are a linear concept.* But, as shown by the example of Space, Dimensions can be multiplexed to create higher order composites, like Planes and Spaces, that have different properties than simple Lines.

*Figure 10. Mental Extents underlie all others*

Certain other Dimensions also overlap or intersect to form recognizable composite situations, though the relations that connect them are not strictly geometric. These include the dimensional complexes of space-time, the temporal-physical, mind-body, individual-social, socio-linguistic, and the material-spiritual. Note how most composites encompass classic categorical oppositions.
Each dimension serves to classify the kinds of things that exist according to the situations in which humans experience them. So every dimension has a prototypical example for each top-level category. For example, there is (generally at least) one prototype Entity, Property, State, Power, Action and Outcome for each of the 12 Dimensions. Prototypes for each category are identified through a discursive process of repeatedly answering the question: “What is a Mathematical Entity?” each time substituting the next Dimension (See Figures 9, 11 and 12).
Any dimensional prototype from any top-level *Category of Things* may have children or related concepts. *As a condition for inclusion in this schema, the child concept must be set apart from its parent by possessing at least one property that makes it somehow distinct.* Alternatively, binary property pairs may be chosen (i.e. individual-group or part-whole) and instances matching both of the resulting descriptions are named.
For example; A whole, individual mathematical Entity is a *Number*; a whole group mathematical Entity is a *Set*. A part of an individual mathematical Entity is a *fraction* or *decimal*; a part of a group mathematical Entity is an *Element*.

### 4.6.2 Ranges

A Range is a special case of a dimension. It is the conceptual Extent of a given Aspect. (An Aspect is a prototype of a Requisite perspectival Entity. See Section 4.6.6 *Requisites*.) Properties and Relations are both instances of Aspects; though in every day description of Entities, we draw more extensively upon Properties than Relations. As a rule of practice, Properties describe Entities while Relations situate Entities. Ranges are linear representations of a given Aspects’ relation to at least one among a number of related properties.

The conception of Range described here is a device used to sort through the relations that hold between properties. Given a way to structure the relations between properties, we can eventually begin to make inferences about the things that possess these properties. The initial premise is that every property falls somewhere on a Range. At either extremity of this conceptual extent or continuum there are two properties. At one end lies the ‘perfect’ example (prototype) of the property in question; at the other end lies the property which describes its ultimate (binary) opposition. Any number of other properties may also be defined as falling between them (e.g. between black and white are innumerable shades of grey.)

Classically, having a given property implies not having certain others. Alternatively, in this conception, when ascribing a Property to any Thing, you are really
indicating the degree to which the thing possesses that Property by situating that thing on the Range usually associated with the given property. The degree to which the thing possesses the property is inversely related to its possession of certain other properties on that range. So, describing something as 100% black means it is also 0% white (Kosko, 1993). Midway (50%/50%) is not neither-nor, but rather both-and; where this range includes a perfectly balanced grey at the midpoint between white and black. A Property, then, is a value in degrees between two or more prototype points on an aspectual Range, where the measure of the extent to which a thing possesses either property, or any of the properties between them is given as its degree. Descriptions of its degree can range in precision from numerical measurement to pragmatic estimates.

While the Range is conceived as the linear relationship of any property to (at least one among) a number of related properties, Ranges can be extended into planes, spaces, (Guarino et al., 2004) or N sized hypercubes by including the inverse (binary opposition) of any interposing properties on a given range to create a second, related intersecting range. In an alternate conception, the relation defining the two properties on a Range does not have to be a binary opposition. 49

4.6.3 Modalities

Modalities are ways of being that prototypically situate States. A Modality effectively marks the existential status of any given thing. Since the relations that exist between the top level categories allow for the deduction of states from a given property, Modalities are in effect meta-states or existential states that imply meta-properties. I have identified two binary pairs of existential states (Realities–Unrealities and Possibilities–

49 It could be Complimentarity, Synonymy, Meronymy, etc.
Impossibilities) and arranged them so as to overlap. This arrangement produces the following continuum: Realities – Possibilities – Unrealities – Impossibilities. Each of these four primary Modalities may have related or children conceptions. For example, 

Currency, Actuality, Factuality, Probability are instances of Realities that can be stratified by Dimension. Whereas most Modalities are prototypically associated with one Dimension, some can be associated with two or more Dimensions. As used in this schema, Modalities situate States existentially.

Existential modalities could also be extended to include the truth values of textual statements. However, as a general rule of practice, this schema assumes that the process of meaning creation between author and reader is collaborative rather than an adversarial. Consequently, authorial statements are not usually subjected to stringent logical proofing.

---

50 Colors highlighting each Instance express prototypical Dimension. See Fig.8 Dimensions of Existence.
or existential inquest. Rather, much as in typical conversational practice, interlocutors are assumed to have some degree of trust towards each other. This does not mean that errors and corrections are impossible; authors may certainly be mistaken, erroneous, and illogical. However, the initial use for this schema is for the automated analysis of texts. Identifying mistakes, errors, and various deviations of logic remains the province of the reader. So while the truth values of statements are not initially necessary, the schema can scale to accommodate other sufficiently meta-level modalities.

The concept of Modalities can also be extended to situate Actions. In this sense, Modalities are manners, ways, or methods of doing or acting; given conditions of functioning, or status. In language, this role is taken up by adverbs, which are verb inflections that express how the action or state is conceived by the speaker. However, we must be cautious here, since we have already identified Properties as describing Entities (including actions). For the purposes of this schema, the modalities of Actions are generalized as Properties.
For example, **Currency** is prototypically a temporal Reality that refers to the present. According to its dictionary definitions, Currency is an external, certain, objective, and verifiable present Reality. **Factuality** is a past temporal reality that is true, genuine, and verifiable. **Probability** is a future temporal reality that is external, uncertain, likely, and objective.
4.6.4 Extents

Extents are the furthest point to which something can be opened, stretched, straightened out, or exerted. They can be conceived of as limits of capacity that situate Powers. But recall that Dimensions measure Extents. Accordingly, an ‘Extent’ names the thing underlying the Dimension whose existence calls for measurement. Dimensions demark the manageable parts of Extents by describing the properties and relations of things that belong there. As a raw conceptual abstraction, Extents are an imposition of limits upon a linear conception of the infinite, the indefinite, or the eternal. For example, they can be conceived as existing from the point at which the infinite becomes finite to the point beyond which the infinitesimal recedes from view.

However, all things have extents and not all such limits are entirely arbitrary. The extents of many physical objects pertain to its size and range of function or duration. For example, the most obvious material extents of an umbrella are its physical surfaces and edges, perhaps its surface qualities, and its mass. Its functional extent could refer to either its broadest or smallest possible configurations when opened and shut. Its spatial extents likely refers to its size and location, whereas its temporal extent could be either the actual duration of its usefulness, an abstract projected ‘shelf life’, or its age.

Things inherit their own extents by prototypically belonging along an Extent (and hence to a Dimension). These individual extents of things are catalogued by Dimension (Figure 11). Consequently, when measured, these extents (Size, Range, Duration, Surfaces, Edges, and Mass) become recognized as Properties of things. Before this, they are the indefinite Extents of things; known to apply, but not yet precisely described.
### Figure 15. Extents of Dimensions

<table>
<thead>
<tr>
<th>Extents</th>
<th>Magnitudes</th>
<th>Distances</th>
<th>Surfaces</th>
<th>Capacity</th>
<th>Durations</th>
<th>Incidences</th>
<th>Sizes</th>
<th>Reaches</th>
<th>Scopes</th>
<th>Capacities</th>
<th>Civilizations</th>
<th>Vocabularies</th>
<th>Understandings</th>
</tr>
</thead>
<tbody>
<tr>
<td>mathematical</td>
<td>Degrees</td>
<td>Lengths</td>
<td>Areas</td>
<td>Expanses</td>
<td>Moments</td>
<td>Instances</td>
<td>Proportions</td>
<td>Spans</td>
<td>Ranges</td>
<td>Memory</td>
<td>Subcultures</td>
<td>Glossaries</td>
<td>Focuses</td>
</tr>
<tr>
<td>linear</td>
<td>Amplitude</td>
<td>Widths</td>
<td>Stretches</td>
<td>Volumes</td>
<td>Periods</td>
<td>Frequencies</td>
<td>Levels</td>
<td>Ambits</td>
<td>Purviews</td>
<td>Intelligences</td>
<td>Cultures</td>
<td>Lexicons</td>
<td>Comprehensions</td>
</tr>
<tr>
<td>planar</td>
<td></td>
<td>Breadths</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>spatial</td>
<td></td>
<td>Heights</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>temporal</td>
<td></td>
<td>Depths</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>temporal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>physical</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>corporeal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>perspectival</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mental</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>social</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>linguistic</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>contextual</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
4.6.5 Orders

It is not a simple exercise to conceive of the prototypical example of an abstract universal like Order. When we try to describe such abstract Entities, we invariably end up describing our own conception of the Concept in question. Recall that, for the purposes of this paper, a Concept is a group of lexically related words that belong to the six Categories of Things described in Section 4.5 Existential Categories. As such, an Order can be both an (abstract) Entity and a property, state, power, action, or outcome of any entity. For example, in systems dynamics and complexity theory, certain systems are described as ordered. But as a property, order is not universal, since not all things express a property of order. Some are disordered or distinctly chaotic. So the inclusion of a Orders as an Element of Context should not be misconstrued as a naïve conception of existence as completely ordered. Order is simply included as a requisite of the human capacity to identify, recognize, and categorize things along contextual Dimensions.

Orders have existential implications for Dimensions, since a concept of Order must exist in order for any system of measurement to exist. Prototypical empirical systems of measurement typically involve a demarcation of boundaries at standardized intervals to create individual units of measurement. Any such progressive demarcation of elements is a simple kind of order. Yet while Orders are an existential requisite for Dimensions, Orders are also typified by Dimensions, in that prototypical instances of Orders can be identified for each Dimension.

So every Dimension has prototypical instances of Order that serve to situate entities within that dimension. In this schema, Orders are arrangements of elements that prototypically situate Actions. This is an obvious association since, more than any other
kind of thing, Actions can be described at successive moments in time as being at
completely different steps or stages. We conceive of Actions as ordered composites of
intermediate steps where each step is situated by the preceding step. Nevertheless, Orders
can serve to situate any Entities.

Figure 16. Related Instances of Orders by Dimension

<table>
<thead>
<tr>
<th>Orders</th>
<th>Logics</th>
<th>Principles</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mental</td>
<td></td>
</tr>
<tr>
<td>math.</td>
<td>Ordinals</td>
<td>Modes</td>
</tr>
<tr>
<td></td>
<td>Numbers</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Num erics</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Progressions</td>
<td>Propositions</td>
</tr>
<tr>
<td></td>
<td>Increments</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Decrements</td>
<td>Premises</td>
</tr>
<tr>
<td>linear</td>
<td>Alignments</td>
<td>Conclusions</td>
</tr>
<tr>
<td></td>
<td>Columns</td>
<td>Arguments</td>
</tr>
<tr>
<td></td>
<td>Rows</td>
<td></td>
</tr>
<tr>
<td>planar</td>
<td>Arrays</td>
<td>Methods</td>
</tr>
<tr>
<td></td>
<td>Matrices</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Diagrams</td>
<td></td>
</tr>
<tr>
<td>spatial</td>
<td>Arrangements</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Placements</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Distributions</td>
<td></td>
</tr>
<tr>
<td>temporal</td>
<td>Series</td>
<td>temporal</td>
</tr>
<tr>
<td></td>
<td>Episodes</td>
<td>Chronologies</td>
</tr>
<tr>
<td></td>
<td>Sequences</td>
<td></td>
</tr>
<tr>
<td>temporal</td>
<td>Chronologies</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Dates</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Lists</td>
<td></td>
</tr>
<tr>
<td>physical</td>
<td>Compositions</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Components</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Composites</td>
<td></td>
</tr>
<tr>
<td>corporeal</td>
<td>Configurations</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Figures</td>
<td>Figures</td>
</tr>
<tr>
<td></td>
<td>Gestalts</td>
<td>Gestalts</td>
</tr>
<tr>
<td>actional</td>
<td>Processes</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Steps</td>
<td>Steps</td>
</tr>
<tr>
<td></td>
<td>Stages</td>
<td>Stages</td>
</tr>
<tr>
<td>actional</td>
<td>Functions</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Procedures</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Operations</td>
<td></td>
</tr>
<tr>
<td>perspectival</td>
<td>Correspondences</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fidelities</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Approximates</td>
<td></td>
</tr>
<tr>
<td>perspectival</td>
<td>Patterns</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Symmetries</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Schemas</td>
<td></td>
</tr>
</tbody>
</table>

121
4.6.6 Requisites

Requisites are prototypically existential requirements that situate Outcomes. This means that every Outcome that occurs has some Requisite that made it possible. However, since all Entities are Outcomes; all Entities have Requisites (See Section 4.5 Existential Categories and Figure 7 Phenomenon of Human Experience).

The concept of Requisite plays an important role in the schema by situating Categories with respect to each other. Recall, at the top categorical level, that Actions are Requisites of Outcomes; Powers are Requisites of Actions; States are Requisites of Powers; Properties are Requisites of States; and Entities are Requisites of Properties. Each Category of Thing from Entities onward is a requisite for the next.

This means that Outcomes don’t happen without Actions; nor do Actions take place without Powers; nor do Powers exist without States; nor do States transpire without Properties; nor do Properties manifest without Entities. If you like, there is an existential cascade from the assumed existence of real Things (that we call Entities) down through the remaining Categories of Things. Of course, by describing all Entities as Outcomes, we hint at a notion whose strongest form is Causation, but that can be adequately described in a more diffuse way as Requirement.

More broadly, Requisites express the intuition of generic necessity. They can effectively re-label any Entity with a meta-level relation that serves to connect it with other Entities. In general, for Entities, Properties, States, Powers, and Actions, this creates an expansive web of connections, some of which may be incidental implications, but which can nevertheless be mined to reveal important structural features of the relations between concepts.
One way to map and follow this web of implications is to start by classifying an Entity by its relevant experiential Dimension. Take, for example, Location. WordNet’s first entry defines location as *a point or extent in space* (WordNet 2.1, 2006). Recall that the experiential Dimensions are a device used to stratify things according to the situation in which their existence is experienced. The definition quite clearly situates Location ‘in space,’ so it is reasonable to conclude a Location is prototypically a spatial kind of Entity. This definition also suggests that a location is a kind of point or extent. So Points and Extents must exist; and they must be the sort of things that have instances (of which Location is an example). In the final analysis of this brief first part of the definition, Spaces, Points, Extents, and (implicitly) Situations are Requisites of Locations.

Of course, each of these requisites have been established as Entities or even Categories in this schema, so there is no need to further explore the definitions of each of these terms. But if you did, you would find an expanding web of concepts that are necessary to understanding these terms. And there still remains more entries in WordNet and other dictionaries’ definitions of Location to explore. While I will not continue this lexical exploration any further in this paper, in order to build and populate this schema, I have in fact used this process to situate the variations of roughly 3,000 terms as Instances of the *Categories of Things*, and of the *Elements of Context* that serve to situate them.

I had originally labelled this category as ‘Primitives,’ thinking atomistically that for every Dimension, there are certain Things that are utterly and inescapably

---

51 The associated Categories of Things can be easily populated by looking at a dictionary to find different accepted forms of the word, or by just imagining them. Accordingly, locative, locational, located, locating, locatable are Properties; locatedness is its State; locativity is a Power; locate is an Action; and location is an Outcome.
foundational and in some important sense, indivisible. But there is nothing especially ‘primitive’ about these things. Some are actually quite evolved or complex concepts. In any case, the concept of Requirement evoked the necessity of relations between concepts.

Their complexity surfaces when looking at the definition of a prototypical spatial Requisite. Bounds are a line or plane that indicates the extent or limit of something. The concept of Bounds requires that Lines, Planes, Indicators, Extents, Limits, and Things exist. Again, most of these concepts have been previously established in the schema. But from this definition, we also find that Extents have Indicators, and that Lines and Planes are Bounds if they serve as Indicators of Extents. This conditional definition hinges on the Actions that things take, or the roles that they play.

Before I can recognize the Extents of a given thing, Bounds must exist (and I must be able to recognize their signature Indications). Given the pivotal role of Extents to Dimensions, and of Dimensions to this classificational schema, whatever is required to recognize Extents is indispensable to the schema. Originally, I believed that since my attempt to operationalize Context relied so centrally on Dimensions, I would have to substantiate both their existence and that of any of the things on which Dimensions depended for their existence.

Consequently, Figure 17, *Dimensional Requisites* enumerates the concepts that the definitions of Extents, Measures, and Magnitudes rely upon. Since you could theoretically expand the scope of the existential dependencies of definitional terms infinitely, Requisites were the last tier, or the furthest I was willing to go to justify the concept of Dimensions.
**How To Read Figure 13**

**Positions** are linear Requisites. A particular Position is a *Point*; An aggregate Position in motion is a *Path* or *Line*.

**Orientations** are planar Requisites. An extrinsic Orientation is a *Direction*; an intrinsic Orientation is an *Axis*.

**Bounds** are spatial Requisites. A linear Bound is an *Extremity*; a planar Bound is a *Perimeter*; a spatial Bound is a *Surface*.

**Continuums** are temporal Requisites. A incomplete Continuum is a *Part*. A complete Continuum is a *Whole*.

**Scales** are physical Requisites. *Units* are particular Divisions of Scales. *Standards* are acknowledged Criteria of Scales.

**Acts** are corporeal Requisites. *Beginnings* are initial Points of Acts. *Completions* are final Cessations of Acts.

**Aspects** are perspectival Requisites. *Properties* are intrinsic Aspects of Entities. *Relations* are extrinsic Aspects between Entities.

**Comparisons** are mental Requisites. *Differences* are distinguished Points of Comparison. *Similarities* are equivalent Points of Comparisons.

**Collectives** are social Requisites. *Persons* are particular social Collectives. *Cultures* are aggregate social Collectives.

**Representations** are linguistic Requisites. *Signs* are specific linguistic Representations. *Symbols* are generic linguistic Representations.

**References** are contextual Requisites. *Significances* are inferred References of Signs. *Denotations* are the particular References of Symbols.

---

**Figure 17. Dimensional Requisites**

<table>
<thead>
<tr>
<th>Requisites</th>
<th>mathematical</th>
<th>linear</th>
<th>planar</th>
<th>spatial</th>
<th>temporal</th>
<th>physical</th>
<th>corporeal</th>
<th>perspectival</th>
<th>mental</th>
<th>social</th>
<th>linguistic</th>
<th>contextual</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cardinals</td>
<td>Quantities</td>
<td>Rates</td>
<td>Points</td>
<td>Paths</td>
<td>Directions</td>
<td>Axes</td>
<td>Units</td>
<td>Standards</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positions</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Extremities</td>
<td>Perimeters</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Orientations</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bounds</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Continuums</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scales</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Acts</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aspects</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comparisons</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Collectives</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Representations</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>References</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
5.1 Schematic Review

Any program that automatically classifies texts from any discipline must have a classificational schema. Rather than expect the policy analyst to come up with one that can accommodate texts from any discipline, I propose the following schema be implemented and tested in software that can automatically classify texts and portions of texts for policy analysts. Accordingly, the purpose of this schema was to identify a parsimonious (Werth, 1999g) set of Categories of Things and related Elements of Context that can eventually be implemented to automatically analyse digital texts. Accordingly, I have proposed specific relations between Categories of Things (Entities, Properties, States, Powers, Actions, Outcomes) that can facilitate cross-categorical inferences. These categories are also related to grammatical parts-of-speech, such that their instances can be identified consequent to the identification of their part-of-speech.

The Elements of Context (Dimensions, Ranges, Modalities, Extents, Orders, and Requisites) have both a prototypical scope (each serves to situate a specific Category of Thing) and broader application (each can situate any Category of Thing). These categories of context can ultimately be considered meta-properties that necessarily belong
to all things. Nevertheless, not all things mentioned in texts explicitly instantiate these contextual meta-properties, nor are they required to.\(^5^2\)

The first three Elements of Context (Dimensions, Ranges, Modalities) are conceived of conceptual devices whose primary relations form a circilinear path. They are intended to delimit context in a useful and intuitive manner. While the raw concept of a Dimension is not innately intuitive, the idea of identifying the different kinds of situation in which humans experience things is understandable.\(^5^3\) The concept of the Range attempts to highlight relations extant between properties, as well as the existential implications associated with the possession of any given property. The Modality slot explicitly acknowledges that existence constitutes the primary (assumed) state of being of all (real) things, and intimates further that all states of being are continuously modified by the possession of properties.

The last three Elements of Context (Extents, Orders, Requisites) were first conceived of as placeholders to organize what I originally identified as the existential prerequisites of Dimensions. Their instances were determined by growing a semantic

---

\(^5^2\) That is to say, even if the range of an entities’ properties are not directly specified, or its modality is not clearly stated, or just because potential scope is not delimited, or an action is not mapped out procedurally, or an outcomes’ antecedents are not identified, this does not mean that instances belonging in these Categories of Things do not have these meta-properties. Rather, these situations just mean that extant meta-properties are sometimes inferred or inherited from previously established Elements of Context.

Moreover, an Entity’s prototypical association with one Dimension does not preclude it from belonging to others. The relevant implication of the linguistic phenomenon of polysemy is that a single form of word can effectively evoke different senses in different contexts of usage. This schema can accommodate the complexity of reference (Halliday et al., 1976) (when paired with part-of-speech tagging) by assigning a term to its proper Category of Things based initially on its functional part of speech. Eventually, after the various facets of a Concept have been categorized by Dimension (and after its Properties have been ranged, its States modified, its Powers enunciated, Actions ordered, and its Outcomes situated) the Contexts of usage for a Concepts’ various senses will be captured and catalogued. So rather than being forced to search by a single term or phrase; users will be able to search for a precise Instances in a Category of Thing by the relevant Elements of Context. For example, a search for ‘Network evaluation’ could be limited to the social Dimension. Then the search could be refined by excluding certain combinations of the search terms.

\(^5^3\) While the term Dimensions could reasonably be expanded into ‘Experiential Domains’ for the purposes of colloquial explanation, in practice the modified terminology would become untenable when discussing the Experiential Domain of Mathematics or Mind.
network of definitions, initially of all terms evoked by the definition of Dimensions and subsequently by the definitions of all terms evoked in their definitions and so forth, in an exponentially growing web of words (See Section 3.6.6 Requisites). For this reason, the instances shown in figures are labelled as the Extents, Orders, and Requisites of Dimensions. However, this does not imply that Dimensions are the only things that these categories situate.

The logic of the schema explains how this identification of the Extents, Orders, and Requisites of Dimensions is both reasonable and necessary: Dimensions are Entities, and as such, they possess Properties. Given that the Elements of Context can each situate any type of thing; it is not unreasonable to identify how certain Elements of Context serve to situate the primary Element of Context. To put it more plainly; since Extents, Orders and Requisites are Elements of Context, they are supposed to situate things. Since Dimensions are things, they, too, are situated by having Extents, Orders, and Requisites.

The schema includes six Categories of Things that are premised upon the Principle of Existence and six Elements of Context that are premised upon the Principle of Situation. It identifies twelve Dimensions that stratify all things according to the situation in which they are experienced. It develops Ranges as a conceptual device that exploits and explodes the intuition of paired oppositions to situate an infinite number of Properties in relation with one another. It proposes four primary existential Modalities and offers a mechanism for accommodating other Modal classes. It names the prototypical Extents along which the properties of things become known as they are measured. It identifies prototypical instances of Orders that serve to situate things by
Dimension. Finally, it identifies the prototypical Requisites of Dimensions as a class of thing that is necessary to this schema of things and their context.

The schema is arranged with one slot for each Category of Thing, each with a corresponding meta-slot for its prototypical Element of Context. Instances of a given Category of Thing are initially classified as prototypically belonging to one of 12 experiential Dimensions. (Subsequently, as required, its relevant Extents are measured, its properties are identified and situated, its existential Modalities are identified, its Orders outlined, and its Requisites mapped.)

Dimensionally prototypical instances may have children instances, usually considered synonyms, but which are in fact distinct from their parent in their possession of at least one distinguishing property. Wherever possible, the corresponding binary opposite of any property is identified (i.e. concrete for abstract; individual for group; particular for aggregate), to determine an appropriate pair for the solitary child instance (See Figure 19).

**Figure 18. Binary-paired Properties serve to structure Child Instances of Categories**

<table>
<thead>
<tr>
<th>Aspects are Requisites of the perspectival dimension.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aspects are intrinsic Aspects of Entities.</td>
</tr>
<tr>
<td>Traits are subtle, abstract Properties.</td>
</tr>
<tr>
<td>Features are obvious, concrete Properties.</td>
</tr>
<tr>
<td>Characteristics are particular Features.</td>
</tr>
<tr>
<td>Qualities are innate Characteristics.</td>
</tr>
<tr>
<td>Principles are basic, essential Qualities.</td>
</tr>
<tr>
<td>Natures are essential Qualities.</td>
</tr>
<tr>
<td>Essences are concentrated, inherent Natures.</td>
</tr>
<tr>
<td>Dispositions are attitudinal Qualities.</td>
</tr>
<tr>
<td>Temperaments are emotional Qualities.</td>
</tr>
<tr>
<td>Attributes are ascribed Characteristics.</td>
</tr>
<tr>
<td>Identities are aggregate Features.</td>
</tr>
<tr>
<td>Modes are functional Properties.</td>
</tr>
<tr>
<td>Manners are behavioural Properties.</td>
</tr>
<tr>
<td>Relations are extrinsic Aspects between Entities</td>
</tr>
</tbody>
</table>
A visual depiction of this structure (See Figure 18) makes use of nested levels of hierarchy to describe the relations between concepts. In classical ontologies, hierarchies evoke increased specificity at deeper levels and increased generality at higher levels, such that the top levels evince the greatest possible level of generality and inclusivity. In some cases, this has the perspectival effect of reducing the perceived importance of items lower in the hierarchy.

However, at some other location in this schema, each Concept in the depicted hierarchy is a ‘top level’ Instance that can potentially have its own children. Allowing a Concept to appear in the schema as both a ‘top level’ Instance and as a child instance of any number of other dimensionally prototypical Instances means that a deeply nested hierarchical structure can be described with only one level of child instances. This removes the constraint of potentially having an infinite number of multi-generational children, allowing the schema to present a manageable depth. (A nested hierarchical view can be automatically generated by the software from the data base as required.)

5.2 Chapter Overview

I began this thesis with a practical intention, an analytical task, and a theoretical motivation. I wanted to help policy analysts by evaluating tools and recommending improvements tailored to their work. My thesis was that most tools would not prove suited to the policy analysts’ operational requirements. Accordingly, when existing tools do not meet the needs of users, the bases of their failure should be examined in the light of theoretical and applied knowledge from relevant domains in order to refine and tailor tools to the needs of users.
In Chapter One, I described the rationale for my self-directed and individually-designed course of study by Special Arrangements in Public Policy and Knowledge Management. In Chapter Two, I evaluated and recommended textual analytical tools and personal knowledge management practices to help policy analysts in their daily work. I established that many currently available tools would not be ideal for the policy analyst and would require extensive modification to merit an unqualified recommendation for adoption. In Chapter Three, I described what ontologies are and how they can be used to improve textual analysis and knowledge management tools. In Chapter Four, I developed a classificational schema to accommodate ‘things and their context’ for future implementation in automated textual analysis software.

5.3 Future Research

In Section 1.3, Ideal Formulation, I outline the desired functionality for a collection of textual analysis tools premised on theoretical and applied work in linguistics that could be used to test and refine theories of textual and discourse meaning creation in readers. In Section 2.5.9 Natural Language Processing, I identified one software tool, VisualText, that can be used to build textual analysis software. In Section 2.6.1 Best Fit, I describe the subset of functions that would be most useful to policy analysts. In Section 2.6.3 Practical Recommendations, I describe how the capacity to classify texts from any disciplines requires a particular kind of ontology. In Chapter 4 What does the Schema Offer?, I develop an ontology intended for implementation in an automated textual analysis tool. In the future, I intend to build a prototype text analysis tool that implements this schema to automatically categorize interdisciplinary texts. I hope to do this supported by a trans-disciplinary PhD program.
APPENDICES

Appendix A: Technological Futures

The system I envision prioritizes ease-of-use and core functionality over immediate and even short-term feasibility.\(^5\) Except where explicitly indicated otherwise, all the hardware technologies that I envision pressing into the practical service of knowledge work already exist. While to some my vision may sounds too much like science fiction, the only fiction I readily acknowledge is that technologies have reached the apex of their usability; that they cannot be designed to be any more adaptive to human capacities and needs. Not all technological innovations are sufficiently user-friendly. Until both hardware and software technologies attain a basic standard of usability, even where they do currently exist, they cannot be considered ready for implementing in the field of knowledge work.

The analyst of the future will have a ‘multi-modal’ interface with their computer work station. Multi-modal interfaces extend the range of sensory feedback channels that are possible with a computer. The traditional modalities of human-computer interfaces (HCI) include tactile and visual movement. Symbolic tactile motions (keystrokes and mouse movements) have long been our primary means of interacting with the computer; and representational visual motion (a screen image refreshed at a constant rate) has been the computer’s primary means of indicating a response. Recent

\(^5\) In the first 50 years of computing, several extraordinary paradigm shifts succeed in created enormous markets. For example, though the 60s the president of IBM insisted there would never be a market for personal computers in the home, now we know better. I expect that the evolution of hardware will continue and within 5 to 10 years, as the devices I describe become affordable, they will also be universal.
modal extensions include both iconic (sound effects) and syntactic (music) audio feedback to augment the computer's visual display.

Old display technologies projected a flat perspective using a (more-or-less) flat cathode ray tube (CRT) screen. The traditional metaphors for the interactive space displayed on screen was appropriately drawn from flat, matrix-like things like 'documents,' 'folders,' 'windows' and 'desktops.' In previous times, this constrained interaction of symbolic tactile motion to moving a mouse on a flat space. However, incredible increases in computational power have lead to ever higher degrees of verisimilitude in the visualization of virtual 3D environments. Even now, the traditional desktop metaphor is evolving into a fuller three dimensional metaphor (i.e. Almost as if in homage Apple's faux 3D, high-gloss, semi-transparent 'Aqua' look and feel; the new 3D Windows 'Aero' user interface features centrally in the new Windows Vista operating system, the successor to Windows XP.)

Nevertheless, the prevailing display technology of today still visualizes three-dimensional scenes from the flat perspective of a photographic or video camera, effectively on a flat LCD or CRT screen. However, 3D screens that create the illusion of depth are currently on the market, as are interactive holographic projections.

New display technologies have resurfaced the problem of facilitating human interaction with virtual 3D objects projected in real space. While so-called immersive 'virtual reality' systems have long existed that surround the user (with goggles) with a 'full' simulated visual field and track user motion with sensors worn on the body that translate motion into symbolic bodily-kinaesthetic movement in the virtual world, these systems were often clunky and intrusive. Their extraordinary cost suited them for
specialist application in niche markets like industrial or military training, engineering, molecular modelling, a few other domains of scientific visualization, and lately, for serious gaming. Progress in the development of VR interfaces suggests that further refinement of spatial tracking technologies will lead to integrated sensors embedded in fabrics and jewellery that will effectively render real three dimensional motion in virtual worlds. This will make the systems somewhat less intrusive.

Other approaches to tracking 3D motion use optical cameras to observe the user from different angles to control interactions with holographic projections. Analysis of eye or head motion captured on camera has also been used to control mouse movement in traditional, flat display computer interfaces.

With the addition of the three-dimensional motion of our bodies, the specific motion of our eyes, the tone of and content of our voices, as well as the introduction of a fuller range of tactile interaction and feedback, human-computer interaction will almost covers the full somatosensory range. The new HCI modalities extend the tactile, visual, and audio channels to now receive and deliver feedback both ways. The repertoire of tactile interaction is extended to include digital pen and paper, touch screens, and resistive force-feedback. (Yes, synthetic smells are also available, though their usefulness to knowledge work is dubious at best.)

The ‘final’ step is direct neuronal interface, which has been successfully negotiated via implanted electrodes with paraplegic patients and in neuro-controlled robotic prosthetics. Alternatively, non-invasive devices can control computer performance using only the mind using electro-encephalitic graphing (EEG) hardware. External, ambient sensors of bodily functions are the most likely trajectory, though
British Telecom has been working toward micro-scale implantable sensation-storing hard drives since the late eighties. Shy of having our personal computers grafted into our nervous systems while in utero, we can expect mainstream human-computer interfaces to cover a wider range of human sensory modalities. Sensors to translate motion into computer commands may eventually supersede hardware keyboards and mice. Miniaturized computer displays will be embedded in traditional eyewear. Immersive, interactive holographic projections will virtualize our whole bodies. 3D displays will observe and virtualize our eye motion. Sensors of metabolic, cardiovascular, and neuronal activity will coach and redirect our mental focus and physical state into target zones. We will speak to our computers and listen to them.

While some of these ‘advances’ may seem extreme, unwarranted, or invasive, in principle only those technologies which are socially sanctioned will reach the public market. This, however, does not mean that these technologies will not be available to advanced technologist institutions, like the military and intelligence communities.

The notion of a ‘work-station’ will be replaced by a personal area network (PAN) and a personal computing network (PCN). This change is already taking place, with the various flavours of wireless connectivity on the market today. The principle difference between today’s status quo and the future will be that the traditional workstation, a desktop tower, computer monitor, keyboard and mouse, will be superseded by any-place access to personal computing resources. Individual consumers will purchase powerful home-based computing resources that help monitor and manage household needs, but which have dedicated wireless connections for family members, who can
connect to home from anywhere using the components of their personal area network as a multi-modal interface.
Appendix B: Example Policy Brief

Problem Definition
1. *Too few policy analysts use knowledge management (KM) strategies, techniques, and tools (STTs).*
2. Too few students of policy analysis know how to use KM STTs.
3. MPP does not teach is students how to use KM STTs.

Variable Identification

Too few = only some of the very youngest.

*Policy analysts* = persons trained or experienced in the process of policy analysis.

*Use* = apply quickly and effectively in daily analytical routines.

*Knowledge Management* = a formalized approach to organizing and mobilizing information and knowledge at the individual and organizational levels.

*Strategies* = systematic approaches to information seeking and processing activities.

*Techniques* = methods of information seeking and processing.

*Tools* = technologies, both hardware and software, for information seeking and processing.

Symptoms

- High diversity of information.
- High volumes of information.
- Low awareness of access to *personal* tacit knowledge (procedural ‘know-how’)
- Low and difficult access to *group* tacit knowledge (sharing ‘know-how’)
- High cost of experimentation.
- High cost of adaptation to specific individual or organizational context.
- Low general awareness of available KM STTs.
- Low understanding of potential value added.
- Low org. buy-in to the values of KM.
- Low org. capacity for organizational change.
- Low org. commitment to implementing KM STTs.
- High cost of implementing KM STTs is borne by analyst.

Context

On any given day, the average policy analyst must handle incredible volumes of many kinds of information. Moreover, they must be able to quickly and effectively marshal their own knowledge and that of their team toward the timely production of analytical resources for their decision-making clients. While Knowledge Management offers ‘best practice’ techniques, tools, and aids to the productive management of these resources, they are not widely used among most policy analysts.

The average (aged 45 to 55 +) policy analyst is extremely unlikely to use digital resources as compared to young (aged 25 to 45) policy analysts (Sorian et al., 2002). As KM STTs are best suited for use with digital resources, it is fair to conclude that the average policy analyst who relies on paper systems will not use these STTs.
However, on average, the student of policy analysis is likely to be both younger and more comfortable using digital informational sources. Also, students are more likely to be approaching their work with an overtly teachable frame of reference, and may be more willing to absorb the cost of learning how to use KM STTs effectively if the immediate and long-term individual benefits are apparent.

If they can be shown to add value to the analytical routines of policy analysis (and to the extent that better policy analysis leads to better policy decisions and improved social outcomes) then it is the public interest to train students of policy analysis how to use KM STTs effectively. In this regard, professional training programs like Simon Fraser University’s Masters of Public Policy program which do not offer their students training in the STTs of knowledge management are failing to serve the public interest to the best of their capacity.

Goals
- Improve access to high-quality policy-relevant information for policy analysts.

Objectives
- Demonstrate the value added offered by integration of KM STTs into PA routines to individual analysts and to their organizations.
- Increase the proportion of policy analysts who regularly and actively use KM STTs.

Alternatives
1. Status Quo
2. Test and Rank KM STTs along the following criteria
3. Develop and Pilot as a KM STT Workshop for MPP students
4. Teach KM STTs as a part of MPP Research Methods Curriculum

Criteria
(1=For ranking Alternatives; 1-3=For ranking KM STTs)

1. Cost of Technology (Upfront, Ongoing)
   a. Cost in $
   b. Cost in Time
2. Ease of Use (Soft-, Hardware)
   a. Degree of Complexity
   b. Degree of Effectiveness
3. Feasibility of Alternative (Upfront, Ongoing)
   a. Degree of Usefulness
   b. Frequency of Use

Impacts
The US dollar cost of alternatives is negligible where existing software and hardware can be used and if new technologies are only purchased when the above objectives cannot be met using current technologies. Considerations for Alternative 4 include a Workshop Instructor salary or the time cost in salary (1/17th) for the existing Research Methods Instructor and possibly purchasing Software licenses for students and staff.

The time cost (of seeking, finding, and retrieving information; processing and determining its policy-relevance; organizing, managing, and re-purposing its content) is measured in minutes required using a given tool in order to produce 1 hour of useful work product. Reduction in time cost is the principle decision rule for selecting KM technologies for implementation.

Additionally, however, the technology’s usability and implementation feasibility over time are considered. Measures of usability include perceived degree of complexity where ‘10’ is most complex; and number of hardware or software failures in 100 attempted uses. Measures of feasibility include perceived degree of usefulness where ‘10’ is most useful; and frequency of actual usage.

Trade-offs
In order to facilitate the use of most KM tools, documents must be digitized. The class of KM software that accomplishes this task is Optical Character Recognition software. OCR software takes scanned images of documents or PDFs and converts them into editable text files (Word documents).

For marking up document contents in a structured and searchable system, there is a class of KM software typically used for Qualitative Research Management. Alternative software for PC were compared against each other and a MAC equivalent was identified.
For managing bibliographic information of cited works and for inserting and automatically formatting and managing citations in reports or documents written in Word, the class of KM software is called Reference Management Software. Software for PC was compared and a MAC equivalent was identified.

<table>
<thead>
<tr>
<th>Citation Management Software</th>
<th>Alternatives</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cost of Technology (US) Cost in $</td>
<td>CM Software</td>
</tr>
<tr>
<td>$110</td>
<td>$110</td>
</tr>
<tr>
<td>($/hr) Cost in Time Software</td>
<td>120</td>
</tr>
<tr>
<td>Degree of Complexity (/10)</td>
<td>5</td>
</tr>
<tr>
<td>Rate of Failure (/100)</td>
<td>2</td>
</tr>
<tr>
<td>Degree of Usefulness (/10) Feasibility of Implementation</td>
<td>3</td>
</tr>
<tr>
<td>Frequency of Use (/wk)</td>
<td></td>
</tr>
</tbody>
</table>

For automated indexing and summarization of documents, the class of KM software is called Content Analysis or Natural Language Processing software. No evaluation, comparison or recommendation is currently available for this class of KM software.

**Recommendations**

- Select Alternatives 2 and 3 in order to further evaluate the feasibility of Alternative 4.
- Across all classes of KM software, use the free Demo software for the first month, in order to facilitate KM STT Workshop.
- After the first month:
  1. **OCR Software**: Purchase OmniPage Pro 15 and use ReadIris Pro 8 for free.
  2. **QRM Software**: Purchase Atlas.TI 5; Evaluate NVivo 11 when available.
  3. **RM Software**: Use RefWorks for free unless offline work is required; then, purchase Reference Manager 11.
  4. **CA/NLP Software**: TBD

**Explanations**

1. Identify useful KM STTs and tailor them to the policy analysis context.
2. Develop and pilot a KM Workshop offered to current (1st and 2nd year) and former (practicing) MPP students. Include in the Workshop a control group of experienced (older) policy analysts.
3. Evaluate feedback from Workshop participants to ascertain how the learning experience ranked on the above criteria.
Appendix C: Determining a Binary Opposite

It is not clear that a term’s syntactic part of speech provides sufficient guidance for which type of binary it should take. (This is a feature of natural language usage and understanding that defies strict rule-based logic. In order to explain it, we must consider an explanation of how language works that does not rely exclusively on a syntactic perspective.)

There are other avenues available for determining the binary opposition of a given term. We could try applying logic of set theory to the question. The problem can be brought to a finer point by thinking with Venn type diagrams. Take the property ‘happy,’ represented by a circle in the diagram. All around it is the void that includes all other ‘non-happy’ properties. But the antonym term ‘unhappy’ is unique. It is a part of, though distinct from most of the larger set of ‘non-happy’ properties. As the opposite of ‘happy,’ it is not only generically ‘non-happy;’ it is also particularly ‘unhappy.’

To recap, a negation contains less information than an antonym. It is more generic. All it tells us is that it is not-X. It could be anything else. An antonym is more precise, because it effectively filters out all other values from universe of possible ‘non-’ properties. So which is the correct binary opposition? The non- or the un-? If precision is the decision criteria, then the correct binary opposition for a term is its antonym. Accordingly, the binary opposition of material is immaterial, not ‘non-material.’

What about terms that either have either no antonym, or have an antonym that does not share a morphological root with the original term? Terms with a morphologically unrelated antonym will still take that ‘different-rooted’ antonym as their
binary opposition. But the morphologically unrelated binary term is related to its original term only through its relationship with the original terms’ negation. (i.e. ‘concrete’ maps to ‘abstract’ through its relationship with ‘un-concrete.’ Without its lexical relation to ‘non-concrete,’ ‘abstract’ would have an indeterminate relation with ‘concrete’.)

For terms with a morphologically related antonym (with prefixes un-, in-, im-, a-, an-, anti-, etc.), that antonym is directly related to the original term. Finally, for terms with no lexical antonym, morphologically derived or otherwise, the negation of the original term stands in as a substitute for a correct binary opposition.

So, while the accepted binary opposition of matter is non-matter, this is not a true binary opposition. Having come to this conclusion, we are still left with the fact that antonyms are lexically derived. The implication here is that antonyms are defined as much by a lexicographers’ determination of correct usage as by logic derivation. And we are still left with a need to determine which of the semantic (synonyms, homonyms, hypernyms) and semiotic (colloquial and disciplinary senses) senses for the synonyms of accepted antonym binary oppositions are acceptable as synonyms of the original terms’ antonyms.
Appendix D: Smart Practices for Managing a Large Textual Collection

The analyst has a pressing need to organize and analyse the digital documents they face in course of their daily work. As mentioned in Section 1.3 *Ideal Formulation*, though I cannot offer policy analysts a tailored suite of tools, I can offer some practical suggestions for implementing personal knowledge management using the available tools. Some are offered in bullet form. I describe in detail steps for creating a sensible file system, and for a collection process for gathering and organizing digital texts and preparing them for analysis them with the available tools.

**Customize My Documents**

The first step is to create a sensible filing system. Short folder names are better. Custom icons can help you find your way to your files quickly (Card et al., 1999). I have one folder for every major kind of ‘document’ that I collect, regardless of how many different file formats those kinds of documents may arrive in.
To keep only the most recently used folders on top: Set the Folder View to ‘Tiles.’ Sort the folder view by ‘Date Modified.’ Arrange Icons by ‘Show in Groups.’

To have access to My Documents quickly whenever the taskbar is visible: Drag My Documents onto your Quick Launch Toolbar.

To access your most important or most often used document types quickly whenever the Desktop is visible: Right click on your important folders in My Documents, and Send Shortcut to Desktop.

To make sure your Desktop stays tidy: Place your Recycle Bin in the top Left corner. All new documents saved to the Desktop will always appear right below the trash. Place your
shortcuts to your most frequently used folders in the second row, and you will give yourself a neat column of 8 to 10 documents on your Desktop before it starts to look messy. When it does, it’s time to file. Filing 8 to 10 documents at a time is manageable.

Figure 20. Experiment with Folder Views to show only the Folders you have recently used

Manage Collection

This section describes the ‘smart practice’ I follow for managing my collection of citable documents. It involves 38 steps between finding an article online and having it ready to analyse using the textual analysis tool of choice. Clearly, 38 steps is about 37 too many, but this is as good as it gets. The first 8 have to do with saving the document using a particular naming scheme, downloading its bibliographic information into RefMan, and filing it to indicate its status. The next thirty have to do with making sure the document is
converted into the right formats for analysis. I will only discuss the first 8 in detail (See Figure 23 for a pictorial overview of the subsequent steps).

The first thing to do with a document you like, is to save it to your desktop.

Make sure that you consistently use a file-naming strategy, to ensure that you can find the document later. I recommend a format that mimics the inline citation of your preferred citation style (Author Last Name. YYYY. Title of Article). Including the article or chapter title can provide useful context, for when you have multiple documents by the same Author. This strategy can reduce your dependency on any third party search tools when you need to find your document again. Even in a folder full of documents, you can find it by sorting the folder contents by File Name.
Next, find the article on Google Scholar. Adjust Google Scholar preferences to Download Citations to RefMan. When you find your article, click on the ‘Import into RefMan’ link. RefMan will ask you to select the appropriate Database to import into.

Figure 21. Make sure the Citation Information is Accurate

Once your article is in RefMan, find the new citation in the ‘Imported’ tab. Change the Reprint Status to ‘In file,’ make sure the Ref Type is correct, and correct any typos. If you corrected the Reference Type and the citation actually belongs in another citation database, copy the Citation into the correct database, then delete the entry from the first database. (Though this process may seem unwieldy, since RefMan has a non-standard interface, its not possible to cut and paste citations between databases.)
Once your RefMan citation data is entered correctly, drag the article from the Desktop into the ‘in RefMan’ folder in eArticles. If you do not already have eArticles open, you can drag the article onto the eArticles shortcut, then click on the shortcut to open the eArticles folder. Drag and drop the article onto the ‘in RefMan’ folder.

**Figure 22. Drag the Article from the Desktop to the my documents\eArticles\in RefMan folder**

Notice that there are 18 folders to accommodate the different file types that eArticles arrive in.
Figure 23. Follow the shortcuts and instructions. When complete, go to next step.

If your textual analysis tools need another format, follow eArticles\PDF image\To Do.
Customize Start Menu

- Re-organize your ‘Start’ menu ‘Installed Programs’ folder for easy access to your most important tools.

Create Macros

- Write macros to automate repetitive tasks.
- Give yourself easy access to tools with keyboard shortcuts.

Desktop Search

Use Google desktop, Copernic desktop search, Windows Desktop Search or OSX Spotlight to find your files. You might think that desktop search programs would obsolesce the need to maintain a sensible filing system. However, these search programs have limitations that make a good filing system useful. You should embed the terms you are likely to use in future searches for the document in the document (Authors Last Name
or the Article Title). This is most easily accomplished in the file name, though meta-data fields could also be used. File location, when in brief and aptly name folders, can also provide contextual data that can be used to filter search results.
REFERENCE LIST


154


Thompson, Clive (2001). The Know-It-All Machine: An audacious quest to teach a computer common sense -- one fact at a time. *Lingua Franca, 11*.


