ABSTRACT

Title of Dissertation: DETERMINISTIC INDUCTIVE LOGIC: A MULTI-VALUED LOGIC FOR REASONING ABOUT CATEGORIES

Allen Eddy Brewer, Doctor of Philosophy, 2000

Dissertation directed by: Claude E. Walston
College of Information Studies

In logic, induction is used to formulate theories (generalizations) from specifics, while deduction is used to derive theories (specializations) from axioms (generalizations). Induction has traditionally employed probabilistic models to formulate representations (models) analogous to logic's domains of discourse, where the resulting models are inherently non-deterministic. An implementation of classical deductive logic that applies to organizing conceptual spaces is ontology, where the classification of a universe of discourse may contain multiple levels of hierarchy to support successively refined deterministic subsets, which are used to facilitate conceptual differentiation. In this paper *deterministic inductive logic* is defined and described as the complement of classical (deterministic) deductive logic. The logic described is a multi-valued logic that

supports formulating theories (generalizations) by combining "specifics" into categories or classes. Three primitive operators are defined and described that support: (1) COMBINE(), (2) COMPARE() and (3) CONTRAST(). Examples are provided to show that deterministic inductive logic: (1) provides a method for building classifications by generalizing about specifics either by defining a class from its members, or a super-class from its member classes; and (2) facilitates the creation of hierarchical structures compatible with classical deductive logic (hierarchical subsumption). An approach is also described for implementing deterministic inductive logic to build information structures to organize and manage information analogously to the controlled vocabulary approach to organizing information.

DETERMINISTIC INDUCTIVE LOGIC: A MULTI-VALUED LOGIC FOR REASONING ABOUT CATEGORIES

By

Allen Eddy Brewer

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
Of the requirements for the degree of
Doctor of Philosophy
2000

Advisory Committee:

Professor Claude E. Walston, Chair
Professor Eileen G. Abels
Professor Robert B. Allen
Professor Bernard A. Fusaro
Professor Denny Gulick

TABLE OF CONTENTS

LIST OF TABLES vii

LIST OF FIGURES xi

ABBREVIATIONS xiii

Chapter I – Introduction 1

Research Objective 1

Background 1

Justification for the research 7

Personalization 8

System evaluation 9

Computational efficiency 11

Chapter II Intellectual foundations: review of relevant literature 13

Information studies 13

Information seeking and information needs 15

Relevance 17

Evaluation 19

Intellectual and physical access 24

Summary of information studies issues 33

Logic 39

Probabilistic inductive logic 39

Deductive logic 42

Chapter III - Deterministic Inductive Logic 44
Determinism 45
Inductive classification 64
Enumeration features 66
Generalized combine operator for multi-faceted features 67
Comparing and contrasting feature vectors 71
Negation 75
Basic feature-vector types 77
Discriminants 78
Testing for hierarchical subsumption 80
Testing feature vectors for equivalence 85
Chapter IV – Formulating a vocabulary with deterministic inductive logic 88
Vocabulary based systems 88
Hierarchy 93
Recall, precision, discrimination and fallout 98
Encoding structured information in feature vectors 105
Chapter V – Representational sufficiency in deterministic inductive logic 110
Chapter VI. Application illustration: Document assembly problem 120
Overview 120
Defining feature vectors to represent provisions and clauses in the acquisition context 127
Accommodating the Definitions clause 129

Accommodating a small set of solicitation provisions 134

Defining feature vectors to represent contracting officer preferences authorized in the acquisition context 145

Chapter VII – Relationships between deterministic inductive logic and deductive logic 153

Set Interactions 156

Completeness 159

Proposition interactions 161

Chapter VIII Discussion, conclusions and future research 162

Discussion 162

Novelty claims 164

Implementation claims 166

Multi-faceted classification example 167

Natural language applications 169

Structured information 169

Conceptual clustering 171

Conclusions 174

Future research 176

Appendix A – Index to FAR Provisions and Clauses referenced in Chapter VI 182

Appendix B – FAR Provisions and Clauses referenced in Chapter VI 185

52.202-1 Definitions. 185

52.214-1 Solicitation Definitions--Sealed Bidding. 190

52.214-2 [Reserved] 190

52.214-3 Amendments to Invitations for Bids. 190

52.214-4 False Statements in Bids. 191

52.214-5 Submission of Bids. 191

52.214-6 Explanation to Prospective Bidders. 192

52.214-7 Late Submissions, Modifications, and Withdrawals of Bids. 193

52.214-8 [Reserved] 194

52.214-9 Failure to Submit Bid. 194

52.214-10 Contract Award--Sealed Bidding.. 195

52.214-11 [Reserved] 196

52.214-12 Preparation of Bids. 196

52.214-13 Telegraphic Bids. 197

52.214-14 Place of Performance--Sealed Bidding. 198

52.214-15 Period for Acceptance of Bids. 198

52.214-16 Minimum Bid Acceptance Period. 199

52.214-17 [Reserved] 200

52.214-18 Preparation of Bids--Construction. 200

52.214-19 Contract Award--Sealed Bidding-- Construction. 201

52.214-20 Bid Samples. 202

14.202-4 Bid samples. 202

52.214-21 Descriptive Literature. 205

52.214-22 Evaluation of Bids for Multiple Awards. 207

52.214-23 Late Submissions, Modifications, Revisions, and Withdrawals of Technical Proposals under Two-Step Sealed Bidding. 207
52.214-24 Multiple Technical Proposals. 209
52.214-25 Step Two of Two-Step Sealed Bidding. 210
52.214-26 Audit and Records--Sealed Bidding. 210
52.214-27 Price Reduction for Defective Cost or Pricing Data--Modifications--Sealed Bidding. 212
15.403-4 Requiring cost or pricing data (10 U.S.C. 2306a and 41 U.S.C. 254b). 212
52.214-28 Subcontractor Cost or Pricing Data-- Modifications--Sealed Bidding. 216
15.403-4 Requiring cost or pricing data (10 U.S.C. 2306a and 41 U.S.C. 254b). 217
52.214-29 Order of Precedence--Sealed Bidding. 219
52.214-30 Annual Representations and Certifications--Sealed Bidding. 220
52.214-31 Facsimile Bids. 221
52.214-32--52.214-33 [Reserved] 223
52.214-34 Submission of Offers in the English Language. 223
52.214-35 Submission of Offers in U.S. Currency. 227
BIBLIOGRAPHY 233

LIST OF TABLES

Table 1. Markov Model of Two Dice 42

Table 2. Automobile Model Feature Set 50

Table 3. Auto Example Buyer Preferences 51

Table 4. Probabilities of Buyer Preferences 52

Table 5. Truth-Values Required to Describe Aggregates 54

Table 6. Truth-Values for Individuals and Aggregates 54

Table 7. Deterministic Inductive Logic Truth-Values 55

Table 8. Relationships of Known and Unknown Truth-Values 56

Table 9. Representing Aggregates that Include Unknown Values 56

Table 10. Truth-Value Combinations 58

Table 11. Group Feature Vector Defined by Buyers Logan and Bonnie 58

Table 12. Buyer Preference Probability Model 61

Table 13. Reduced Logan-Bonnie Model 62

Table 14. Market Segment Preferences Defined By Gerald And Allen 63

Table 15. Adding Features by Enumeration 66

Table 16. Multi-Faceted Features 68

Table 17. Combine for the DNA Four-Facet Base-Pair Model 69

Table 18. Compare Operator 72

Table 19. Contrast Operator 72

Table 20. Negation for Binary Features 75

Table 21. Negation for *Indeterminate* and *Unknown* Facets in *n-Facet* Features. 75

Table 22. Negations of Truth-Values in Compare and Contrast Operators.......... 76

Table 23. Logan and Bonnie Combination Vector.. 79

Table 24. COMPARE(COMBINE*(case-list), identify vector*) Truth Table................. 80

Table 25. COMPARE(COMBINE*(case-list), difference-vector*) Truth Table............ 80

Table 26. Gerald-Bonnie Combination.. 82

Table 27. Gerald-Bonnie-Logan Combination 82

Table 28. Test for Inclusion of Logan in the Gerald-Bonnie Segment 82

Table 29. Defining Broader Terms by Combination.. 101

Table 30. Indexing Toys Example 1.. 103

Table 31. Indexing Toys Example 2.. 103

Table 32. Extending Feature Vectors with Unknown Truth-Values.................. 104

Table 33. Representing a Teddy Bear with Extended Features 105

Table 34. Examples of Mapping Quantitative Conditionals to Logical Features107

Table 35. Mapping Quantitative Values to Faceted Features 108

Table 36. Indexing Toys with Limited Features 113

Table 37. Indexing Toys after Extending the Feature Set 114

Table 38. Object configurations.. 115

Table 39. Differentiating Different Configurations of the Objects in Table 38 . 116

Table 40. Identifying Relevant Discriminants (Features 1-6)............................ 117

Table 41. Identifying Relevant Discriminants (Features 7-10).......................... 118

Table 42. Organizing the Features into a Sufficient Set of Facets..................... 118

Table 43. Selected Object Configurations 119

Table 44. Representation Required to Discriminate Objects 119

Table 45. Features to Represent Types of Instruments..................................... 128

Table 46. Additional Features Required by the Definitions Clause................... 131

Table 47. Representation of Definitions Clause.. 132

Table 48. Mandatory Versus Required when Applicable Features.................... 134

Table 49. Encoding Provisions in Figure 21... 137

Table 50. Encoding of Additional Provisions at 14.201-6 140

Table 51. Features Required by Provisions at 14.201-6.................................... 141

Table 52. Provision Encoding Using the Features Identified in Table 53.......... 141

Table 53. Hypothetical Acquisition Conditions.. 142

Table 54. Included and Excluded Provisions Given the Hypothetical Acquisition Conditions in Table 53... 144

Table 55. Features for Controlling Optional Inclusion of Value Engineering.... 147

Table 56. CO Option Encoding.. 148

Table 57. Default Authority - Default Contract Action..................................... 149

Table 58. Superior Authority Exclusion - Default Contract Action................... 149

Table 59. Superior Authorizes Inclusion - Explicit Exclusion by the CO.......... 149

Table 60. Superior Authorizes Inclusion - CO Elects Inclusion........................ 150

Table 61. Precedence Relationships.. 151

Table 62. Mapping Boolean Expressions to Deterministic Inductive Logic...... 154

Table 63. Mapping Compound Expressions – Contradictions 155

Table 64. Mapping Compound Expressions – Subsumption............................. 155

Table 65. Mapping Compound Expressions – Incomplete Representation........ 155

Table 66. Set Interactions in Figure 23 Mapped to Boolean Expressions.......... 158

Table 67. Approaches for Handling Enumerated Sets 160

Table 68. DIL Feature Vector 173

LIST OF FIGURES

Figure 1. Feature vector for Gerald's auto preference 51

Figure 2. Filter model based on Logan and Bonnie preferences.......................... 59

Figure 3. Classification based on Logan and Bonnie’s preferences..................... 60

Figure 4. Classification based on Gerald and Allen’s preferences....................... 64

Figure 5. Logan-Bonnie market segment inductive classification 65

Figure 6. Combine function pseudocode (*n-facet* version)................................. 70

Figure 7. Compare function pseudocode 73

Figure 8. Contrast function pseudocode 74

Figure 9. Generalization formula 77

Figure 10. Formula for creating identity vectors.. 78

Figure 11. Formula for creating difference vectors.. 78

Figure 12. Formula for creating an inclusionary discriminate vector 79

Figure 13. Formula for creating an exclusionary discriminate vector.................. 80

Figure 14. Test for hierarchical subsumption 81

Figure 15. Set relationship subsumption tests... 83

Figure 16. Class relationships formed inductively from individuals 84

Figure 17. Degree of similarity function 86

Figure 18. Degree of difference function.. 87

Figure 19. Levels of contracting authority... 124

Figure 20. Definitions clause 130

Figure 21. Provisions identified by 14.201-6 136

Figure 22. Applicability conditions for Value Engineering 146

Figure 23. Regions defined by the interactions of three sets 157

ABBREVIATIONS

~	logical NOT symbol
*	multiplication symbol
/	division symbol
^	exponentiation symbol
BT	broader term
C	name of a computer language
CO	Contracting Officer
DNA	Deoxyribonucleic Acid
DDC	Dewey Decimal Classification
DIL	Deterministic Inductive Logic
DOD	Degree of difference
DOS	Degree of similarity
ERIC	Educational Resources Information Clearinghouse
ET	Equivalent term
FAR	Federal Acquisition Regulations
FLETC	Federal Law Enforcement Training Center
FORTRAN	Formula Translation Language
HCI	Human Computer Interaction
IFB	Invitations for Bids
IIF	If and only if

IRS	Internal Revenue Service
LCC	Library of Congress Classification
LCSH	Library of Congress Subject Headings
LSI	Latent Semantic Indexing
MOA	Memorandum of Agreement
NEC	not elsewhere classified
NT	narrower term
RT	related term
SDI	Selective Dissemination of Information
SMART	Salton’s Magical Retriever of Text
SNP	Single nucleotide polymorphism
SIFT	Stanford Information Filtering Tool
ST	synonymous term
WAIS	Wide Area Information System

Chapter I – Introduction

Research Objective

The objective of this research is to define and describe *deterministic inductive logic* (DIL). DIL is proposed to become the bottom-up complement of deductive logic, where bottom-up is to be understood as referring to the approach used to formulate generalizations. To qualify as inductive a model must form generalizations from specifics.

The development of DIL results in a multi-valued logic usable for reasoning about aggregates or collections of individuals such as might be characteristic of categories of objects or entities. Functionally, *deterministic inductive logic*:

1. provides a method for building classifications by generalizing about the specifics of the members of a class, and
2. results in a conceptual structure in which the assumptions of hierarchical subsumption, characteristic of deductive logic, are supported.

Background

The two primary approaches to logic are inductive and deductive. Deduction supports the implication of conclusions based upon general principles, axioms or laws of nature, while induction supports the formulation of generalizations from specifics such as observations or objects. Deductive models are reliant on a system of *a priori* axioms. Inductive models require no *a priori* presumptions.

The presence of *a priori* requirements limits a system's adaptability particularly in any context different from what it was designed to describe. While there is some potential for error in the design of a deductive model, a more important potential source of error exists in the application of deductive models, particularly where there are any potential differences between the design context and the use context. Inductive models present the potential to avoid or mitigate this source of error.

Deductive models have traditionally been based upon an ontology that defines a context. That ontology is defined by all of the axioms and relationships between and among axioms applicable to the context being modeled. A complete theory of a universe of discourse (context) includes all of the theorems that are provable from the interaction of the axioms in that universe, given the rules of logic. A classification is another term for a deductive model or ontology.

Classifications employ hierarchy to support the representation of more than one level of generalization. Classifications may be either monohierarchical or polyhierarchical. In a monohierarchy all subordinate concepts are linked through a single path of the conceptual structure. Monohierarchies can be represented with tree diagrams. A polyhierarchy contains more than a single conceptual hierarchy and is frequently used to model the set of concepts that become possible by combining multiple monohierarchies.

Induction, the formulation of generalizations to reflect some set of specifics, such as empirical evidence, has traditionally been implemented using probabilistic models (such as Markov models). These models are built from cells that contain calculations or estimates of the probability of phenomena or combinations of phenomena. Probability models, while theoretically deterministic for the specific static collection of objects used to determine that model's cell probabilities, are generally deployed and used non-deterministically to estimate probabilities of specific phenomenon or combinations of phenomena. Bayesian models are used to estimate conditional probabilities for phenomena given certain antecedent phenomena (conditions) and are inherently non-deterministic since the process of construction is one of aggregating observations or specifics.

Non-deterministic models are quantitatively based information representations in which quantities are represented by values chosen from a continuous scale within a range of possible values applicable to the specific characteristic, feature, aspect, etc., being represented. A quantity is typically used in a non-deterministic model as a coefficient to represent the relative weight, strength or probability of a feature in a class or in a category that the model seeks to represent.

A deterministic deductive model may employ quantification in first order logic to describe the state of all objects (universal quantifier) or the state of at least one object (existential quantifier) in a collection of objects. Non-deterministic models generally describe the presence of a feature in an object in a collection in terms of the probability that a randomly selected object would exhibit the characteristic of interest using a range of zero (no probability) to 1.0 (universal quantifier) where $0 \leq$ probability ≤ 1.0.

Historically, classification has been conceptualized as a deductive structure, where the top of the hierarchy represents the universe of discourse or universal concept, which subsumes everything of interest for the purposes of the classification. The bottom of the hierarchy is represented by terminal nodes, which are the most specific conclusions, or intentionally rich categories or classes that are of interest. Classification provides structure to information irregardless of whether the structure is used to organize concepts, types of plants, animals or children's toys.

In a classification used to represent a conceptual structure, hierarchy helps define the relationships between concepts. Broader terms or generalizations are hierarchically higher in such structures than are the terms used to represent concepts subsumed by the broader terms. When a concept is divided into facets the narrower concepts assigned represent non-overlapping subsets of the broader set or class.

A classification, such as the Dewey Decimal Classification (DDC) is based upon a design formulated in accordance with Aristotelian principles, in which the universe is successively divided according to a sequence of rules of division or *differentiae*. The Dewey approach to organizing information is implemented by classifying items of a collection in terms of the pre-existing conceptual model. One uses the axioms that regulate the organization of information to identify the place where an item should be located deductively.

In the Library of Congress Classification (LCC), one can observe more easily the effects of continuous improvement. Periodically single categories are reorganized into sub-classifications of smaller categories to reflect the evolution of ideas represented in publications such as the effects of relativity on the conceptual model of physics. This is typically done when a category or class has gotten sufficiently large that the term used to define that class has insufficient pragmatic precision simply because its scope covers more than a desirable range of items. A term that is too general to be practical for retrieval, relative to a collection, can be subdivided to improve the discrimination of aspects contained within the class represented by that term.

Classifications, such as LCC and DCC, need to undergo incremental improvements over time as the empirical contents of the collections they are intended to organize change. Each change results in the classification or conceptual model being refined to reflect evolving relationships between user's needs for information and the presence of items in the collection. The specific contents of a class undergoing subdivision are used to reorganize that class in terms of generalizations subsumed by the overall class. While this process is based loosely upon inductive principles, the process is best characterized as a search for the axioms that regulate deductive division as opposed to a process of bottom-up association of objects based upon their retrieval or use.

A faceted classification is a state-based model formed from a set of qualities (facets) that are (1) mutually exclusive features, aspects or characteristics of the objects or members of collections being represented by the model; and (2) the facet set must be collectively exhaustive. Collective exhaustivity may require completing the facet set with a "not elsewhere classified" facet that can act as a remainder to account for objects that express a characteristic other than the ones identified in the enumerated list of previously observed facets. While this approach insures that all objects can be represented it does not insure that all characteristics of objects can be represented. Adding newly discovered facets requires representational extensibility, which is not possible for deductive models that are inherently *a priori*.

An alternative approach to creating a hierarchical organization arises from defining the base nodes of the hierarchy as individual empirical evidences and conceptualizing the levels of intermediate hierarchical nodes as being formed by the combination of lower tier nodes. Each generalization in such a conceptualization is based upon combining a number of subordinate nodes and inheriting the characteristics of those individuals that are common to the group or class defined by the generalization. Classes of classes can be formed in the same manner as classes of individuals, successively, until at the topmost level all individuals are accounted for by the "universal" which is represented by the top node of a hierarchy. The approach to forming a classification described in this paragraph is inductive.

Justification for the research

No definition or description exists for building a deterministic classification inductively. The principle contribution of this research is to define and describe a multi-valued logic that can be used inductively to discover and describe categories, and because it is deterministic, can also be used deductively for analysis, retrieval and reasoning about individuals and aggregates.

Logic is qualitative. It is designed to represent the presence or absence of qualities. It is intended for the manipulation of statements about individual entities and classes or categories of entities, irrespective of whether the entities are concepts or objects and irrespective of whether the entities have or do not have any provable or demonstrable existential referent.

The specification of an approach for constructing logic-based classifications inductively contributes an alternative to existing methods designed to create inductive models that employ quantitative values chosen from continuous range scales.

The definition and description of a deterministic complement to deductive logic is believed to be a contribution, which may be useful in organizing, managing, analyzing and generally working with information.

Potential benefits of applying this research might arise from its potential to contribute to facilitating:

- Personalization
- System evaluation
- Computational efficiency

Personalization

Classifications are constructed to represent conceptual structures that are to be used by large populations. The cost of building a classification and indexing objects using the classification is sufficiently substantial that it is only economically justified by large-scale use. A common conceptual structure cannot be expected to reflect every possible individual perspective, so some degree of mismatch and inefficiency can be expected to exist where incorporating the representation necessary to support specific individual's retrieval needs cannot be economically justified.

A personalized classification, particularly one designed to structure concepts for an individual, for example for a college student working on a paper about the socio-economic effects of war on rural villages in Viet Nam 1955-1970, is essentially not economically possible because it can not be produced as a byproduct by an information system. A student might construct a classification in the form of an outline, by hand, to organize their research, however, there are no information system products that automatically assist in building such outlines or conceptual structures. If a technology were available that could be used to assist in the construction of a personalized classification or conceptual schema, as a byproduct of accessing and using information, then personalized classifications that reflect the unique viewpoints of individual users could become economically feasible.

The capability to reflect an individual's perspective, viewpoint and interests might become economically possible if a methodology existed that could automatically formulate generalizations by examining user information behaviors particularly when those behaviors include organizing objects into user defined categories.

System evaluation

The two metrics most frequently used for system evaluation are recall and precision, where both are measures of the performance of a system based upon the relationships between an information need, the set of retrieved documents and the contents of the collection from which the set was retrieved. Recall specifically addresses the degree to which all relevant objects were retrieved, while precision addresses the degree to which only relevant objects were retrieved. Both recall and precision are represented by a value from a continuous scale ($0 \leq \text{value} \leq 1.0$) that represents a degree from *not* to *perfect* in terms of a percentage. In vocabulary-based systems, characteristic of deterministic deductive systems, the assignment of terms used to represent categories or classes directly affects both recall and precision. The exhaustivity of indexing interacts with the specificity of indexing in achieving recall and precision. Where there is a mismatch between the representation needs of the user and capabilities of the system, performance, as measured recall and precision, is degraded.[1]

An alternative evaluation approach could be defined in terms of a representation's capability to describe a user's information need both completely and correctly. Where a perfect match exists between a user's retrieval intentions and a system's representation, a correctly formulated request would result in both perfect recall and perfect precision. In such a case the system representation might be termed *sufficient* to describe the user's need. Where there is less than perfect recall or precision, the representation is *not sufficient* either because it is too general (insufficient specificity) or the application of indexing terms is insufficiently exhaustive (every relevant object was not associated with every relevant term).

Viewing a representation in terms of its sufficiency or non-sufficiency with respect to a specific collection, vocabulary, and need is not meaningful for classifications that are designed to be used by large populations, particularly where economics necessitate limitations on representing every possible viewpoint. It is reasonable to expect that every deterministic deductive classification is insufficient with respect to some information need, however, if the classification were built inductively and personalized; one could assess the

[1] Dagobert Soergel, *Organizing Information: Principles of Data Base and Retrieval Systems,* (Orlando, FL: Academic Press, Inc., 1985).

need to extend a representation by defining a metric of *representational sufficiency*.

If deterministic inductive logic supports vocabulary extensibility, a system evaluation metric of *representational sufficiency* might be used to assess when to extend a representation. Representational sufficiency in this case is a binary value (*true* or *false*) that is defined by the sufficiency of a representation to represent a specific information need such that the term correctly selected to represent the information need is one used to conceptually represent a category containing objects, such that the retrieval of all the objects assigned to that category results in both perfect recall and precision relative to the information need and collection.

A methodology that contributes a system evaluation metric for representational sufficiency could provide the capability to assess the need for extending a representation and could contribute a system trigger that could be used to initiate an inductive learning process to tune the system's representation to reflect the information needs of a specific user.

Computational efficiency

Inductive probabilistic models require representations that can account for every possible phenomenon and every potentially relevant combination of possible phenomena. The characteristic of traditional inductive models, because of exhaustivity, tends to result in combinatorial explosions that render large and complex problems computationally intractable. Increasing complexity results in exponential increases in computational requirements unless a model can be constructed that can eliminate factors or partition problems in ways that allow meaningful problems to be solved with small subsets of the information required in a model that represents all possible situations.

If one reduces a model initially to only reflect the homogeneous characteristics of objects of interest the number of possible combinations that are of interest may be able to be substantially reduced in proportion to the degree of natural organization present in a collection. Restated, in highly organized collections there are far fewer relevant characteristics than there are in chaotic or random collections.

A deterministic inductive model that can identify necessary features for discriminating meaningful categories of objects might be useful in reducing the size of a model and might improve the computational efficiency of systems by reducing the number of categories and characteristics in a model's representation to the minimum required to achieve representational sufficiency relative to a specific use.

Chapter II Intellectual foundations: review of relevant literature

The intellectual foundations for understanding the origins, motivation and the potential application of deterministic inductive logic are based upon research commonly associated with information studies. In addition to this research base, there are some fundamental characteristics of inductive and deductive logic that form the basis for defining and describing deterministic inductive logic. This chapter addresses:

- information studies concepts that form a framework for deterministic inductive logic; and
- the fundamental logic issues that are necessary to understand later chapters.

Information studies

The general field of information studies includes many disciplines and aspects of disciplines that address information, its meaning, communication, storage, retrieval, compression, access, use, description, etc. It has traditionally encompassed situations in which humans interacted with information. However, information studies also can include systems that are entirely computer based.

The field of information studies is sufficiently large that authors have ordinarily focused on aspects that are fairly narrowly defined that address concepts appropriate to their interests. Four researchers whose work is relevant to the foundations of deterministic inductive logic, Marchionini,[2] Soergel,[3] Goker[4] and Kuhlthau,[5] each deal with human information access and use, but focus on different issues and approach their research with different perspectives. For example, Marchionini's focus on browsing and serendipity in finding information is very different from Soergel's focus on the analysis and procedures necessary to build systems to organize information so that it can be found analytically. Goker addresses probabilistic induction and the role of machine learning algorithms for imputing user need specifications. Kuhlthau addresses the process and human factors that are evidenced by students engaged in information seeking activities associated with learning objectives. Deterministic inductive logic does not match up exactly with any of these researchers but falls into a middle area that intersects each.

The information studies issues that contribute to placing deterministic inductive logic into some perspective are organized in this chapter into the following four interrelated aspects:

---

[2] Gary Marchionini, *Information Seeking in Electronic Environments*, (New York: Cambridge University Press, 1995).

[3] Soergel, *Organizing Information*.

[4] Ayse Safiye Goker, "An Investigation into the Application of Machine Learning in Information Retrieval" (Ph.D. diss., London: City University, Department of Information Science, 1994).

[5] Carol C. Kuhlthau, "Inside the Search Process: Information Seeking from the User's Perspective," *JASIS*, *42*(1991): 361-371.

- Information seeking and information needs
- Relevance
- Evaluation
- Intellectual and physical access

Following a review of relevant work in each of those four aspects, a summary draws together threads to describe the basis of deterministic inductive logic.

Information seeking and information needs

The concept of information seeking has been extensively studied in terms of human users seeking information to fulfill needs. Taylor[6] defined four types of information need (visceral, conscious, formalized, and compromised) that occurred during the process in which a user evolved from a potentially unrecognized need to the point when they were gathering information and needed to compromise the specification of their underlying need to represent that need in terms of characteristics that could practically result in acquiring information. An information need has been variously described as a gap,[7] an anomaly[8] or a

---

[6] R. S. Taylor, "The Process of Asking Questions," *American Documentation* 13(4) (1962): 391-396.

[7] Brenda Dervin, "Communication Gaps and Inequities: Moving toward a Reconceptualization," in Dervin, Brenda and Voigt, Melvin, eds. *Progress in Communication Sciences* (Norwood: Ablex, 1980), 73-112

recurrent interest, such as for current awareness in a Selective Dissemination of Information (SDI) system.[9] In every case information seeking is a goal oriented learning process, which is based in a human, cognitive context.

Dervin's research has focused on information as a user construct and upon how information seeking is a sense making process in which a user builds an understanding or sense for the problem and information sources and facts that can contribute to understanding and ultimately solving the problem (need) that initiated their search.[10] Kuhlthau observed in the students that she studied that, while many thought they were gathering information throughout the seeking process, most were refining their understanding of the problem or objective during most of the information seeking process and spent time at the end gathering information that was actually used in writing papers.[11]

People make sense of situations by associating and differentiating similar and different objects or concepts, which they organize by comparing and

---

[8] Nicholas J. Belkin, "Anomalous States of Knowledge as a Basis for Information Retrieval," *Canadian Journal of Information Science* 5(1980):133-144.

[9] Douglas W. Oard, "The State of the Art in Text Filtering," *User Modeling and User Adapted Interaction* 7(3) (1997): 141-178.

[10] Brenda Dervin, "Information as a User Construct: The Relevance of Perceived Information Needs to a Synthesis and Interpretation," in *Knowledge Structure and Use: Implications for Synthesis and Interpretation* (Philadelphia: Temple University Press, 1983). 155-183.

[11] Kuhlthau, *Inside the Search Process*.

contrasting facts, observations, concepts, etc. Deterministic inductive logic is intended to be a tool that contributes to organizing conceptual structures in terms of a user's situational frame of reference. It is intended to assist users in defining concepts by associating observations and objects into categories and by identifying the common implicit and explicit characteristics that are expressed by the members of those user defined categories.

Relevance

One can contrast the inexact and developmental human process of information seeking with a more automated or mechanical process such as a known item retrieval that might be characterized by a stored procedure in a database management system designed to retrieve all records from a specific information store with a specific set of criteria. In both contexts, however, the information objects that should be delivered by a system are described by the concept of *relevance* in information studies.

Relevance addresses the appropriateness or fit of an information object to a specific use. Saracevic[12] addressed relevance in terms of the effectiveness of contact between a source and a destination in a communication. Cooper[13] looked

---

[12] Tefko Saracevic, "The Concept of Relevance in Information Science: A Historical Review," in *Introduction to Information Science* (New York: R. R. Bowker Co., 1970).

[13] W. S. Cooper, "A Definition of Relevance for Information Retrieval," *Information Storage and Retrieval* 7(1) (1971): 19-37.

at the logical implication between two concepts in which a concept was relevant if it was part of an inferential chain linking a proposition to a conclusion. Wilson[14] defined relevance as situational, and described relevant information as that information that could result in the change of an individual's views in relation to a question of concern in a situated context. Soergel[15] adopts Cooper and Wilson's views of relevance and differentiates pertinence and utility, where *pertinence* deals with "aboutness," and *utility* addresses an object's potential to contribute to solving or understanding a specific problem. Green's research addresses the limitations of topical relevance[16] and looks at the problems of representing linguistic structures[17] that can contribute to relevance determinations. Wang[18] delved into relevance from a utilitarianism perspective and identified situations in which users engaged in satisficing during their information seeking processes.

---

[14] Patrick Wilson, "Situational Relevance," *Information Storage and Retrieval* 9(8) (1973): 457-471.

[15] Dagobert Soergel, "Indexing and Retrieval Performance: The Logical Evidence," *JASIS* 45(8) (1994): 589-599.

[16] Rebecca Green, *Topical Relevance Relationships. I. Why Topic Matching Fails*, University of Maryland/CLIS, 1994.

[17] Rebecca Green, "The Expression of Syntagmatic Relationships in Frame-Based Indexing," (Ph.D. diss., University of Maryland/CLIS, 1989).

[18] Peiling Wang, "A Cognitive Model of Document Selection of Real Users of Information Retrieval Systems," (Ph.D. diss., University of Maryland/CLIS, 1994).

Deterministic inductive logic seeks to meet Cooper's demands for inductive logical implication, and Wilson's requirements for situational determinants. By using feature vectors, we seek to encompass the types of information structures (frames that contain hierarchical levels with multiple slots) Green employed to refine relevance determinations beyond the topical level.

Evaluation

A user, a subject expert, or some other authority such as a gatekeeper pragmatically judges relevance. The relationship between a query and a document, in a systems context is a common basis for understanding and determining relevance.[19] If one defines relevance, however, in terms of the quality of match between the query and object, the definition of relevance diverges from the theoretical appropriateness of an object to a specific user in a situated information need or use context and becomes more like a basis for evaluating a system's performance. The system performance perspective of relevance, which addresses the match between query and object, is different for controlled vocabulary systems than for vector space and ranked retrieval systems.

Evaluation is ordinarily conceptualized in a system context in which an information need is represented by a query and where a query formulation is used

---

[19] Carlos A. Cuadra and Robert V. Katter, "Opening the Black Box of Relevance," *Journal of Documentation* 23(4) (1967): 291-303.

to extract relevant documents from a collection of documents.[20] The two most common evaluation criteria used to assess the quality of the retrieved collection are *recall* and *precision*.[21] Recall is a measure of completeness, while precision is a measure of correctness. Recall is calculated by dividing the number of relevant documents retrieved by the total number of relevant documents in the collection. Recall values are for practical purposes always fractions of unity. Precision is calculated by dividing the number of relevant documents retrieved by the total number of documents retrieved, so it is also a fraction of unity. Two complementary metrics that are sometimes used in evaluations are *discrimination* and *fallout*.[22] Discrimination addresses the quality of the capability of rejecting non-relevant documents. Discrimination is calculated by dividing the number of irrelevant documents rejected by the total number of irrelevant documents in the collection. Fallout is the complement of discrimination and addresses the degree of noise in a system's results. Fallout is calculated by dividing the total number of irrelevant documents retrieved by the total number of irrelevant documents in the collection.

[20] Cyril Cleverdon, "The Cranfield Tests on Index Language Devices," *ASLIB Proceedings* 19(6), 1967.

[21] Soergel, *Organizing Information*.

[22] Ibid.

In a controlled vocabulary context, Soergel[23] addressed the relationships between indexing languages and performance and identified two conceptual issues that directly impact system performance, *specificity* and *exhaustivity* of indexing. Specificity of indexing, as described for controlled vocabulary systems, addresses the level of generality at which concepts are represented. For example specificity addresses whether the concept *nuclear fusion* is represented at the general level of *physics,* at a more specific level such as *nuclear physics* or at the specific level of *nuclear fusion*. Specificity of indexing directly affects discrimination. Low specificity indicates concepts are not as specific as user's interests, which reduces discrimination and precision. Exhaustivity addresses (1) the number of concepts in a document that are represented in the surrogates searched by the system, and (2) the proportion of all the concepts contained in documents that are represented in the indexing language implemented by a system.

What concepts a system can represent will directly affect a system's performance. Performance is a function of (1) the degree to which the system represents the concepts of interest to a user completely and correctly, and (2) the degree to which the system exhaustively represents every concept in a document, and (3) the correctness of index term assignments in surrogates. Errors in creating document surrogates can be expected to reduce both recall and precision. The

---

[23] Soergel, *Indexing and Retrieval Performance.*

performance effects from specificity and exhaustivity are a function of the user's information requirements or needs.

Soergel differentiates between two types of exhaustivity, *viewpoint exhaustivity* and *importance exhaustivity*.[24] Viewpoint exhaustivity addresses whether all the viewpoints expressed in a document that might be useful in retrieval are represented in the systems indexing language, while importance exhaustivity addresses whether all the concepts or just the important concepts in a document are reflected in document surrogates. Where viewpoint exhaustivity is low, some viewpoints will be missing that could reduce recall. Where importance exhaustivity is high, discrimination may be improved by reducing the number of documents for which the concept is only peripheral.

In the 1970s, with the rise in the availability of computing capabilities, alternative approaches to indexing languages that were based upon representing the words in documents were developed.[25] These approaches have been developed for both information filtering and retrieval applications.[26] In all cases there are limitations to the recall and precision that can be obtained from filtering and retrieval systems, where those limits are a function of the ability to accurately

---

[24] Ibid.

[25] William B. Frakes and Ricardo Baeza-Yates, eds. *Information Retrieval: Data Structures and Algorithms* (Englewood Cliffs, NJ: Prentice Hall, 1992).

[26] Oard, *The State of the Art in Text Filtering*.

reflect the information need in a query and the ability of the system to match the query with a representation of the content of information objects. Even employing inductive machine learning techniques to process relevance feedback and automatically improve query formulations, Goker was unable to materially improve the recall and precision of retrievals over that provided by students initial query formulations.[27]

Performance is therefore a function of the capability to represent an information need (specificity of indexing) and the inclusion of all the applicable terms in the surrogates of information objects (exhaustivity of indexing). To determine if the necessary terms are present in a representation to express an information need, a mechanism for assessing representational sufficiency might be useful. Where a representation is sufficient to express a user's information need in a specific situational context, the quality of the query formulation will not contribute to performance degradation. Where the system's representation is not sufficient to express the user's information needs, either the representation should be modified or one can reasonably predict degraded performance when filtering or retrieval results are obtained.

When a user's need cannot be precisely formulated in a system's representation the user should reasonably expect degraded performance. Where a

---

[27] Goker, *An Investigation into the Application of Machine Learning in Information Retrieval.*

formulation is too general, a user can expect to experience reduced discrimination. When a representation is missing important concepts, the user can expect reduced recall. Clearly, during a learning or sense making process as described in the information seeking section above, the precision of a query formulation will fluctuate with the user's appreciation and understanding of the problem and the information space being used, and performance can similarly be expected to fluctuate.

Intellectual and physical access

Archives are typically organized and stored in relation to some structure of information suppliers or sources. For example, records of adoptable children entering the State of Maryland are associated with the department of health in an adoption file because of a legal requirement to file a health certificate when an adoptable child enters the state. A filing order (organization, organization sub-unit, by date) that is used to control physical access can be made more accessible to users by associating records to topical categories. For example, users could more easily find adoption related records if an index was made available to associate all the different records from every department that has or has had responsibility for relevant records. Finding aids (e.g., indices) facilitate users that are unaware of the organizational unit responsible for a particular record or may not have been aware of every department with relevant records.

Academic and public libraries have generally adopted a subject based classification system to assist in both intellectual and physical access. For example, the shelf order of a public school collection may be based on the DDC system or in the case of an academic research library, the shelf sequence might be based on the LCC system.[28] In both cases the clear delineation between intellectual and physical access is somewhat blurred, because the system allows users to find a physical location and then refine their search by browsing the shelves. In both DDC and LCC subject access is enhanced with a cross reference between a topical index and classification numbers.[29] The Library of Congress Subject Headings (LCSH) is a multi-volume set that links topics to the classification categories that contain related information and to other subject headings.

Some of the early computer-based collections, such as the Educational Resources Information Clearinghouse (ERIC) are reliant solely upon computer based intellectual access tools because users do not have access to the physical documents. In these systems the intellectual access links to an accession number that was originally used to retrieve a document or microform for delivery to a user. These types of systems are totally dependent upon an indexing language and a thesaurus to navigate subject terms that are used to locate information products. Since the only way to navigate the collection is through the index and

---

[28] Bohdan S. Wynar, *Introduction to Cataloging and Classification,* 8th ed., rev. Arlene G. Taylor. (Englewood Co.: Libraries Unlimited, Inc., 1991).

[29] Ibid.

abstracts, important research focused on the design and maintenance of indexing languages and thesauri.[30] To improve access economically and to facilitate retrospective linkages, lead-in terms were provided in thesauri to assist users in finding preferred terms.[31] Studies of users and the terms they expected to use to find information[32] allowed some researchers to bifurcate indexing language definition to differentiate between user-oriented and document-oriented indexing.[33]

An alternative to the analysis and design required to build a user oriented indexing language was to probabilistically associate index terms to topics.[34] In a probabilistic system one can formulate a query in the indexing language and refine the query with relevance feedback quantitatively. This approach provides advantages in handling relevance feedback without relying on users to be proficient in Boolean searching.

---

[30] Dagobert Soergel, *Indexing Languages and Thesauri* (Los Angeles, CA: Melville Publishing Company, 1974).

[31] Ibid.

[32] Raya Fidel and Dagobert Soergel, "Factors Affecting Online Bibliographic Retrieval: a Conceptual Framework for Research," *JASIS 34(3)* (1983): 163-180.

[33] Soergel, *Organizing Information*; Raya Fidel, "User-Centered Indexing," *JASIS* 45(8) (1994): 572-576.

[34] Frakes and Baeza-Yates, eds. *Information Retrieval: Data Structures and Algorithms*.

Full text automated indexing systems,[35] such as Salton's Magical Retriever of Text (SMART), provide searchers with access to the contents of documents. These systems were developed to improve upon the capabilities of systems like ERIC that limited user's to search terms. In these systems a feature vector is populated by the frequencies of terms found in documents, after eliminating low information content terms (i.e., conjunctions, articles, anaphora, etc.). Many of these systems allow users to search for specific terms and terms within a specified distance of each other.[36] SMART and ERIC, while clearly retrieval systems, have both been deployed for use in providing current awareness, a common use of SDI systems.

While information filtering is conceptually based upon a recurrent or long-term interest and information retrieval ordinarily addresses a single need, both retrieval and filtering are conceptually sufficiently similar,[37] that for the purposes of providing a foundation for deterministic inductive logic, no differentiation will be made. In research that addressed information filtering, using both the Boolean

---

[35] G. Salton and M. McGill, *Introduction to Modern Information Retrieval,* (New York: McGraw-Hill, 1983).

[36] Frakes and Baeza-Yates, eds. *Information Retrieval: Data Structures and Algorithms*.

[37] Nicholas J. Belkin and W. Bruce Croft, "Information Filtering and Information Retrieval: Two Sides of the Same Coin?" Co*mmunications of the ACM* 35(12) (1992): 29-38.

approach[38] and a vector space model[39] Stanford ultimately adopted the vector space model to support the Stanford Information Filtering Tool (SIFT).[40] Other researchers in information filtering have used techniques such as Latent Semantic Indexing (LSI)[41] to construct representations to match queries with documents. These and other quantitatively based techniques seek to improve system performance by improving the relationship between the representations of queries and documents. These techniques also frequently return document citations that are ordered or ranked by similarity coefficients, which are interpreted to be the probability a document (citation) is relevant given the query formulation used for retrieval.[42]

The quantitative approaches are at least partially motivated by the belief that an example of a relevant document may be more user-accessible from an intellectual access standpoint than an indexing language used in conjunction with

---

[38] Tak Yan and Hector Garcia-Molina, *Index Structures for Selective Dissemination of Information, STAN-CS-92-1454,* 1993.

[39] Tak Yan and Hector Garcia-Molina, *Index Structures for Information Filtering Under the Vector Space Model, STAN-CS-93-1494,* 1993.

[40] Tak Yan and Hector Garcia-Molina, "SIFT - A Tool for Wide-Area Information Dissemination," in *Proceedings 1995 USENIX Technical Conference*, 177-186.

[41] S. Deerwester, S. T. Dumais, G. W. Furnas, T. R. Landauer, and R. Harshman, "Indexing by Latent Semantic Analysis," *JASIS* 41(6) (1990): 391-407.

[42] Oard, *The State of the Art in Text Filtering*.

a thesaurus. Marchionini observed that the intellectual effort associated with an analytical approach to searching was greater than that associated with selecting objects from a list.[43] He has focused on building more user-friendly systems by looking at how instrumentation in the system's interface, such as starfields with sliders to control filtering what documents are represented in a visual display, can provide dynamic user controlled access to information.[44] This research focuses on improving the Human Computer Interaction (HCI), which facilitates efficient browsing or scanning and helps users evaluate alternatives to find relevant documents. [Relevant documents found by users while browsing may be used as exemplars to define a query for submission to a compatible retrieval system to obtain additional information or to change the ranking of documents.] HCI research addresses conceptualizing information spaces in the process of creating an interface design that naturally assists users in finding potentially relevant information. HCI research seeks to help users avoid the proforma analysis required by an analytical problem solving approach that is inherent in the design and operation of controlled vocabulary systems that require users to perform Boolean searches.

---

[43] Marchionini, *Information Seeking in Electronic Environments*.

[44] Ben Shneiderman, "Dynamic Queries for Visual Information Seeking," *Technical Report of University of Maryland, 1994, CS-TR-3022*.

One problem with both ranked systems and systems that focus on user interactions is that they do not provide an approach that allows the user to assess the degree of recall in situations where a user may have the need for a high recall result. In a study of Wide Area Information System (WAIS), a quantitatively based retrieval system, Marchionini, Barlow and Hill observed that there was a probability of finding relevant documents in the bottom quartile of a ranked collection.[45] This evidence severely limits the feasibility of using ranked systems or relying on systems that focus on user interactions, when high recall is required. For example, selecting the most advantageous supplier in a government acquisition context requires soliciting all potentially responsible bidders or offerors. An undesirable result might be obtained by missing some relevant sources that might provide the most advantageous bid or proposal. Similarly, in a structured situation such as drug interactions, constructing an indexing language and thesaurus such as the *Alcohol and other drugs thesaurus*[46] may be justified.

There are a number of additional approaches researchers have taken to develop and improve conceptual navigation for high recall searches including

---

[45] Gary Marchionini, Diane Barlow, and Linda Hill, "Extending Retrieval Strategies to Networked Environments: Old Ways, New Ways, and a Critical Look at WAIS," *JASIS* 45(8) (1994): 561-564.

[46] National Institute on Alcohol Abuse and Alcoholism and Center for Substance Abuse Prevention, *The Alcohol and Other Drugs Thesaurus: A Guide to Concepts and Terminology in Substance Abuse and Addiction*, 2d ed., Rockville, MD: National Institute on Alcohol Abuse and Alcoholism and Center for Substance Abuse Prevention, 1993.

work based upon conceptual graphs,[47] spreading activation,[48] rule based systems,[49] and Shank's work with scripts and plans.[50] This research is sometimes categorized as artificial intelligence. Generally artificial intelligence methods contribute to analytical capabilities, however, they can also be embedded in interfaces to contribute to dynamic user interactions.

The process of induction when implemented by a clustering approach has tended to produce clusters that are numerically different but generally not ones that are conceptually accessible. Michalski and Stepp[51] sought to mitigate this problem by developing a clustering approach that resulted in a tree built from a

---

[47] John F. Sowa, *Knowledge Representation: Logical, Philosophical, and computational Foundations, preliminary ed. ICCS 94*, College Park, University of Maryland, 1994.

[48] Lisa Rau, "Knowledge Organization and Access in a Conceptual Information System," *Information Processing and Management* 23(4) (1987): 269-283.

[49] Frederick Hayes-Roth, "Rule-Based Systems," *Communications of the ACM* 28(9) (1985): 921-932; Joseph Giarratano and Gary Riley, *Expert Systems: Principles and Programming* (Boston: PWS-KENT Publishing Company, 1989); Steven Pollitt, "CANSEARCH: An Expert Systems Approach to Document Retrieval," *Information Processing & Management* 23(2) (1987): 119-138.

[50] R. C. Schank and R. P. Abelson, "Scripts, Plans and Knowledge," in *Thinking: Readings in Cognitive Science* (Cambridge, England: Cambridge University Press; 1977). 421-432.

[51] Ryszard S. Michalski and Robert E. Stepp, "Learning from Observation: Conceptual Clustering," in *Machine Learning: An Artificial Intelligence Approach.* Ryszard S. Michalski, Jame G. Carbonell and Tom M. Mitchell, eds. (Palo Alto, CA: Tioga Publishing Company, 1983). 331-364.

predetermined set of conceptual categories. Their objective was to automatically construct classifications in which all concepts are circumscribed by a conjunctive concept of object attribute relationships.

Their approach was broken into two distinct components (1) clustering and (2) hierarchical assignment. The clustering goal was integrally related to the configuration of clusters so that the clusters, when organized hierarchically, would minimize concept interactions. The clustering component's goal was to produce a maximally disjoint set of clusters that were aligned with predefined categories. Their approach to cluster creation was iterative. It was controlled by metrics for cluster quality.

Each of their categories was defined by relational statements (selectors). All selectors required define a category were collectively a logical-complex. Logical complexes were used for conceptual category assignment. An event satisfied a logical-complex whenever the values in that event satisfied all the selectors in the logical-complex. At the collection level, a collection of events (cluster) for which there is a logical-complex that is satisfied by the events in the cluster and only by the events in that cluster was termed a set-complex.

The measure of cluster quality is a function of the cluster overlap. Michalski and Stepp adopted an *inter-cluster difference* as a criterion of overlap. *Inter-cluster difference* is conceptually the degree of uniqueness of two clusters based upon their degree of disjointness. Disjointness is determined by subtracting the number of non-overlapping object attributes in two clusters from the total number of selectors in those two clusters.

The object attributes included concepts like tonal range, rhythm, sex of singers, etc. These characteristics included both binary domains (yes, no) and faceted domains, such as the degree of embellishment {0,1,2,3,4,5} or harmonic structure {monophonic, polyphonic}. The attributes were selected and ordered by a subject domain expert.

The approach they adopted for hierarchical assignment begins with the total collection at the top node. The hierarchy building process is controlled by a “continue growth” criterion and stops when the criterion fails. This criterion requires that the fit between the clusters and their descriptions at every level of the hierarchy must be better than at the previous level.

In the example reported they developed a classification for 100 Spanish folk songs. The folk songs were indexed to twenty-two musicological attributes.

Summary of information studies issues

Learning and sense-making are inherently inductive processes in which information is successively accumulated and assimilated to develop a successively greater understanding of a problem space. To best reflect the sense-making, learning, and the evolution of information needs characterized by information seeking behaviors, an inductive approach to system operations would seem most appropriate. For these reasons, a significant amount of research in information studies employs the inductive model in applications that range from building intelligent agents for simulation and control applications[52] to selecting movies that might be of interest.

The major drawback of both controlled vocabulary systems and systems that are based on concept navigation are that the concepts, organization and intellectual access are fixed before the user's information need occurs and are immutable during the information seeking process. For example, while conceptual clustering provides a dynamic way of ordering a collection in terms of concepts, it relies on a predefined conceptual structure that cannot change during a search process.[53]

Pre-coordinated index terms are associated prior to a search or retrieval and are most similar to the classification numbers from DCC or LCC.[54] Such systems develop very slowly and are modified very infrequently due to their size, complexity and cost. This problem of a fixed conceptual schema is mitigated to some degree by mechanisms for dynamically associating indexing terms at retrieval or filtering time. These runtime associations are referred to as post-

---

[52] Paul E. Nielson, "SOAR/IFOR: Intelligent Agents for Air Simulation and Control," *Proceedings of the 1995 Winter Simulation Conference, C. Alexopolos, K. Kang, W. R. Lilegdon, and D. Goldsman, eds.* (1995): 620-625.

[53] Ryszard S. Michalski and Robert E. Stepp, "Learning from Observation: Conceptual Clustering," in *Machine Learning: An Artificial Intelligence Approach.* Ryszard S. Michalski, Jame G. Carbonell and Tom M. Mitchell, eds. (Palo Alto, CA: Tioga Publishing Company, 1983). pp. 331-364.

[54] Wynar, *Introduction to Cataloging and Classification.*

coordination indexing.[55] Faceted classifications, where facets are defined as being mutually exclusive and collectively exhaustive sets of differentiae, can be used in post coordination systems to associate index terms at runtime.[56] Post-coordination queries in controlled vocabulary systems are defined using deductive logic at runtime.[57] These systems, even if they employ automated methods for processing documents and applying index terms are heavily reliant upon lexicographers to develop the conceptual structures and facet schemas and consequently they tend to change slowly if at all. Even user-oriented systems designed after extensive analysis of user needs will be necessarily limited to those needs expressed by users during the analytical phase of system development.

Full text indexing systems, recommender systems, and systems based upon inductive models are the least *a priori* and most runtime oriented approaches available. This has justified their application in a number of filtering type applications like Tapestry,[58] and Group Lens.[59] Associating users on the basis of

---

[55] Ibid.

[56] Soergel, *Organizing Information*.

[57] Ibid.

[58] David Goldberg, David Nichols, Brian M. Oki, and Douglas Terry, "Using Collaborative Filtering to Weave an Information Tapestry," *Communications of the ACM* 35(12) (1992): 61-70.

[59] Paul Resnick, Neophytos Iacovou, Mitesh Suchak, Peter Bergstrom, and John Riedl, "GroupLens: An Open Architecture for Collaborative Filtering of Netnews," *Association of Computing Machinery* (1994): 175-186.

their preferences in conjunction with a vector space model permits relatively continual evolution of information that can be used to associate similar user preference patterns to support recommender type filtering systems.[60]

Essentially, information filtering and retrieval research has evolved into a bifurcated set of approaches that are either heavily dependent upon *a priori* information structures[61] or are extremely dynamic and bereft of any conceptual navigational mechanisms (HCI starfield systems). Controlled vocabulary systems are probably the most demanding of user analytical cognitive effort and they tend to require users to formulate and reformulate queries manually as they develop an understanding of the information that is relevant to their problem space. Dynamic query systems and many quantitatively based approaches provide integrated query reformulation by inductively processing examples that take advantage of relevance feedback by modifying query formulations on the fly or filtering the objects represented in a starfield.

Many researchers, particularly ones that identify themselves with HCI, would likely argue that Boolean retrieval and analytical approaches to searching are not particularly user-friendly, while providing tools to users like dynamic queries may result in their fulfilling information needs in a more user-friendly

---

[60] Oard, *The State of the Art in Text Filtering*.

[61] Dagobert Soergel, *Information Structure Management: A Unified Framework for Indexing and Searching Database, Expert, Information-Retrieval, and Hypermedia Systems (Medford, NJ: Learned Information*, 1994).

way.[62] The user's conceptualization of information associated with a problem space is not likely to develop in the same way serendipitously as it might if the user were confronted with a conceptual schema, as is characteristic of a classified organization. If the goal is conceptual navigation of an information space, quantitative and HCI approaches are limited in their ability to provide conceptual navigational tools to contribute conceptual guideposts to users. If the goal is dynamic intellectual access reconfiguration, the complexity, human cognitive effort, and the cost of evolution and change render controlled vocabulary systems inappropriate.

What motivated this research is the belief that what is actually needed is a mechanism in which a user can define concepts by example and where a system can translate information associations into a logic that operates similarly to a vector space model. The goal is a system in which a user can compare and contrast the concepts they have defined using examples. These comparisons are to be used to assess the relative overlaps or non-overlaps between and among concepts. Ideally, the advantages a hierarchy provides in defining a general concept that subsumes a number of subordinate concepts could be provided to users to assist in their analysis and in conceptualizing an information space. If such a mechanism were sufficiently user-friendly, it might contribute an analytical tool for users that wished to dynamically build conceptual structures,

---

[62] Marchionini, *Information Seeking in Electronic Environments*.

using the learning that ordinarily occurs during the information seeking process. If the mechanism is based upon characteristics that can be identified explicitly in objects or even ones that can be identified during information seeking and associated with objects (annotating implicit characteristics); that mechanism could help a user to assess the homogeneous characteristics of sets used to define concepts. Additionally the implicit and explicit characteristics of objects could be used to determine whether an object is or is not a member of some class (conceptual category) defined by example.

The motivation for defining deterministic inductive logic is based upon the assumption that, if a user is permitted to define the specificity of indexing and the system supports user defined concepts (exhaustivity of indexing) the performance of that system in terms of recall, precision and discrimination might match what is required by a specific user with a situated information need seeking to make sense of an information space.

The hierarchical structure of deterministic inductive logic was modeled on controlled vocabulary systems. The feature vector characteristics that will be described in subsequent chapters were derived from the vector space model. Deterministic inductive logic is an analytical tool that is intended to facilitate a controlled vocabulary’s capability to provide conceptual navigation in an information space without requiring a user to become an expert in logic.

Logic

There are a few very fundamental issues in both inductive and deductive logic that are necessary to understanding later chapters. The inductive logic issues are first treated and then the deductive logic issues are identified.

Probabilistic inductive logic

Deterministic inductive logic is derived from modifying probabilistic induction. Determinism is attained by reducing the continuity of a probability model to a set of discrete deterministic states. More specifically, when applied to representing propositions, probabilistic induction provides for a range of probabilities that may be used to describe the truth of a proposition; where a probability of 0.00 is used to represent non-truth (falsity), a probability of 1.00 is used to represent truth, and the range $0.00 < x < 1.00$ is used to describe different magnitudes, estimates or expectations of the possible truth of a proposition.

A probabilistic inductive model is constructed of a set of features or characteristics[63] that are defined to facilitate representing or describing objects or propositions that are the subject of analysis. Individual cases, such as documents, objects, observations, concepts, and etc., are represented in terms of the truth

---

[63] The terms "feature" and "characteristic" are used interchangeably. Both terms are used to refer to aspects, properties, attributes, etc., of an object, document, proposition, etc., that are used in descriptions, in modeling a universe of discourse, in formulating logical propositions, etc. For the purposes of defining and describing deterministic inductive logic, a feature is anything that can be formulated into a truth-functional logical proposition.

(probability = 1.00) or falsity (probability = 0.00) of the specific characteristics of an individual case. There is no conceptual difference between the formulation of truth functional propositions in truth-functional propositional logic and probabilistic inductive logic to describe individual cases. At the unit level, given an existential referent, a characteristic can only be *true* or *false*.

Probabilistic inductive logic facilitates formulating descriptions of concepts that represent classes of phenomena that might be collectively referenced by a term (identifier) that is used to identify a particular concept of interest. For example, the concept chair can be described in terms of the probabilities of certain implicit and explicit characteristics evident in some population of chairs. Probabilistic inductive logic allows the creation of a generalized description of the characteristics of *chairs* based upon an aggregation of the characteristics of some number of representative chairs. The concept represented by the group can also be used to predict the characteristics of an individual chair that has not yet been experienced in terms of what one might expect if they were to randomly select a chair from the collection of chairs used to formulate the concept *chair* by aggregation.

Probabilistic induction is quantitative. The description of any aggregate is defined in terms of quantitative values for characteristics. The description of an aggregate may also be obtained using a formula[64] to weight the probabilities of the characteristics of a class to reflect the distribution (known or expected) of the characteristics of its members. For example, if three out of five cars that roll off an assembly plant were painted black, then the probability assigned to the characteristic "color--black" would be 0.60.

Multiple characteristics can be represented using feature vectors or arrays constructed of an ordered sequence of individual's characteristic's truth-values. More complex models, such as Markov models can be constructed to represent contingent probabilities. An example Markov model that could be used to represent the theoretical probabilities[65] of rolling two distinguishably different (i.e., differently colored) dice is illustrated in Table 1.

---

[64] The formula used to determine the probability of a class is determined on the basis of a specific situation and the information available. If all individuals are available for analysis, an acceptable formula might be an arithmetic average calculated by the sum of individuals (0 = false, 1 = true) divided by the total number of individuals for each characteristic to be represented.

[65] The theoretical probability of rolling any one of the 36 possible values (2 dice each containing 6 faces per die) is theoretically equal for a set of "fair" dice, where each possible face is equally probable. In an experimental situation each time two dice were rolled, the values on each would be used to locate the cell in the table and that value would be incremented by 1 and the total number of trials would be incremented by 1. Experimentally each cell is equal to the sum of the observations of that phenomena divided by the total number of observations.

Table 1. Markov Model of Two Dice

	1	2	3	4	5	6
1	.02778	.02778	.02778	.02778	.02778	.02778
2	.02778	.02778	.02778	.02778	.02778	.02778
3	.02778	.02778	.02778	.02778	.02778	.02778
4	.02778	.02778	.02778	.02778	.02778	.02778
5	.02778	.02778	.02778	.02778	.02778	.02778
6	.02778	.02778	.02778	.02778	.02778	.02778

In a probabilistic inductive model any characteristic, at the aggregate level, may be more or less *true*, or more or less *false*. This makes it difficult to generalize about a class in terms of the characteristics its members express, particularly when some members exhibit a characteristic and others do not. In the example Markov model of two dice, every possibility is equally probable. While the result provides a basis for predicting the characteristics of randomly selected objects, it doesn’t contribute to reasoning about classes of objects, because to have a valid generalization about a class, the characteristics of all members of that class must be homogenous.

Deductive logic

In deductive logic a universe of discourse is defined as the domain containing everything of interest and under consideration. That universe is defined in terms of axioms, which are always *true*, or a system of axioms (ontology) that provides a framework that can be used to classify and organize everything in the universe.

First order deductive logic allows for propositions to be quantified. The two types of quantification are *universal* and *existential*. A *universally quantified* characteristic is a characteristic that is expressed by every individual contained within the scope of the proposition. An *existentially quantified* characteristic is a characteristic that is present in at least one individual contained within the scope of the proposition.

One can map probabilistic inductive values to a quantified deductive formulation by treating the two probabilistic endpoints 0.00 and 1.00 as two cases of universal quantification and treating the probability range $0.00 < x < 1.00$ as indicative of existential quantification. This mapping results in defining three distinct states on the probabilistic continuum (1) universally *false* (0.00), (2) universally *true* (1.00) and (3) existentially *true* or existentially *false* ($0.00 < x < 1.00$). This mapping forms the basis for deterministic inductive logic.

Chapter III - Deterministic Inductive Logic

The two types of logic we are concerned with in this paper are deductive and inductive logic. Deductive logic is used to test the validity of hypotheses given a system of axioms. In deductive systems one begins with one or more statements that are universally applicable that describe the "universe" in terms of some set of defined concepts. Deduction becomes useful for reasoning by formulating hypotheses in terms of defined concepts and testing for their validity.

The vocabulary or terms used in logic represent qualities. Propositions that have assigned truth-values can be assessed using truth tables. First Order logic allows statements to be quantified. Quantification provides a capability to describe the scope of the applicability of a proposition. A universally quantified proposition is always true. A proposition that is existentially quantified and *true*, is a proposition for which there is at least one instance that is true. The scope indicated by universal or existential quantification is the same irrespective of truth-value. A universally quantified false proposition is never true and always false. Quantified propositions are proved and disproved using rules in deductive logic. For example, De Morgan's rule allows one to exchange the proposition (A and B) with the formulation ~(~A or ~B) by employing negation. The implementation of computer-based "artificial intelligence" that is based on deductive logic requires a fairly sophisticated inference engine capable of modeling and proving theorems using the rules of deductive logic.

Induction, on the other hand, is a methodology for formulating generalizations from specifics. For example, induction is used to combine observations to generalize about the collective experience represented by the sum of those observations. Unlike deduction, where one can prove a hypothetical that might not have been observed by using the rules of logic in a system defined by a set of axioms, induction is limited to generalizing about specifics, typically ones that have been observed. With induction, one demonstrates that a set of observations can be generalized in a certain manner. One cannot prove as in a deductive system anything with induction; therefore it should be seen more as a methodology for associating and organizing information, in this case on the basis of features or characteristics, than as a method for proving hypotheses.

Determinism

One can model a situation by describing it in terms of whatever characteristics are applicable. For our purposes a situation can also be described as a context, which is some portion of a larger physical or conceptual environment or system. If the context is defined as a "marble in a brown cardboard box" the modifiers brown and cardboard that are associated with the box might be understood to be characteristics of the box that holds the marble. Similarly, the marble is a feature of a context containing two objects where the marble is contained in a brown cardboard box.

We will treat the box and the marble in this example context as two concepts. The terms "brown and cardboard" are characteristics of the box. They are also concepts in the described context. If we wished to model the physical situation described, we would define four terms {marble, box, brown, cardboard} and assign truth-values to each concept. By ordering the four concepts, we could formulate a sequence of truth-values such as (TTTF) and (FTTT) which we could associate with the features of the situation {marble, box, brown, cardboard} to represent two different configurations of the concepts. In the first case (TTTF) we would have a situation in which there is a marble and a brown box, while in the second (FTTT) we would have a brown cardboard box but without any marble.

If containment is an important characteristic for understanding the example context, a fifth concept "containment" might be used to associate the marble with the box such that it is either contained or not contained by the box. We could represent (TTTFT) and (FTTTT) in the form of a feature vectors by mapping the ordered sequence of truth values to features, where each feature is associated with a concept or characteristic and where the feature vector (TTTFT) holds the truth-values associated with each of the defined features {marble, box, brown, cardboard, contained-by}.

The interpretation of a feature vector requires one to know what concepts are being represented and in what sequence they are represented. An arbitrary vector such as (TFTTTFFTFFT) is a valid feature vector, however, without knowing what concepts are associated with each truth-value, it is not possible to interpret any specific meaning of, for example, the fifth feature whose truth-value is "T." To be meaningful, a feature vector must be sequenced in a precise order with each feature associated with a concept that can be assigned a truth-value. Features may be characteristics, concepts, objects, or anything that can be assigned a truth-value and interpreted in terms of its associated concept.

Theoretically, a concept for which a quantity could be interpreted, such as the number of male offspring in a family, could be used as a "truth-value" if one allows truth-values to include any number of values that are meaningful for a specific concept. For a multi-valued truth-value to function, the value must be one of a possible set of values that are mutually exclusive and collectively exhaustive.[66] If the set of values is not exhaustive, it is not possible to know what truth-values are possible and hence what the scope of the concept might be. If the truth-values are not mutually exclusive in the way Boolean truth-values {*true*, *false*} are, then it is not possible to distinguish each of the possible different conceptual meanings.

In an inductive system a case is existentially quantified if one can demonstrate that there exists one or more cases that are consistent and complete

---

[66] In library science literature an index term that is selected from a set of index terms that are mutually exclusive and collectively exhaustive is termed a facet. A faceted classification is a deductive tree that is formed by differentiae that are facets. A faceted classification is thus not ordinarily a binary tree unless every facet in the classification has exactly two possible values.

with respect to some set of hypothetical features that would be used to distinguish the type of case under consideration. In the example above, the demonstration that “marbles can sometimes be found in a brown cardboard box” is achieved by finding a case in which there is a marble in a brown cardboard box. The feature vector (TTTTT) with the associated concepts and characteristics {marble, box, brown, cardboard, contained-by} is a demonstration that is sufficiently consistent and complete to justify existentially asserting the existence of “a marble in a brown cardboard box.”

The feature vector (FTTTF) would not demonstrate the “marble in a brown cardboard box” hypothesis because the feature vector, while indicating the presence of a brown cardboard box, shows no indication of a marble and the lack of a marble is consistent with the denial of “contained-by.” The vector (FTTTF) is an existential demonstration of “a brown cardboard box without a marble therein.”

The differentiation between proof and demonstration is very important. Induction is always consistent and complete with respect to whatever has been “experienced.” When one demonstrates an existential hypothesis that hypothesis remains valid so long as the case used to demonstrate the hypothesis remains valid. When one demonstrates a universally quantified hypothesis, however, that hypothesis is subject to review and modification if ever there is a contradictory case that demonstrates that the previously universal concept should become existentially quantified. Since every generalization is derived from “existing evidence” the inductive approach insures that new information can always be accommodated and that as it is accommodated the truth-values can be modified to reflect the “reality.”

This is important both from the perspective of modeling a “universe” and modeling some number of sub-sets of that universe that might be treated as contexts. If a context is only part of the universe of possibilities, it is very possible that while a concept is only existentially quantified at the universal level, it may be universally quantified in some sub-context of the universe. For example, if we were to have a universe of cardboard boxes of many different colors and defined a sub-context that contained all and only brown cardboard boxes, green cardboard boxes, if they existed, would never occur in the brown box context. Tin boxes would not be existentially quantified in a cardboard box context, while tin, cardboard, paper, wood and steel boxes might all exist in some general context of boxes.

The allocation of features in a model depends upon the purpose for which the model will be used. We will create a second example with two concepts each of which has five possible values. The two concepts are style and color and these two concepts will be used to examine how concepts can be used in a deterministic version of inductive logic. The example deals with a set of automobiles in terms of the styles and colors offered by a hypothetical manufacturer. In our example, this hypothetical manufacturer offers five styles in five colors. We will model these two concepts and their respective five possible values by building and manipulating them as an ordered sequence of ten features each of which is represented by a truth-value. We use a one-dimensional array (vector) to store the features. Table 2 defines how the features are ordered in a sequence to form a feature vector.

Table 2. Automobile Model Feature Set

Feature Number	Type of Feature	Feature value
1	Style 1	Civic
2	Style 2	Accord
3	Style 3	Passport
4	Style 4	CRV
5	Style 5	Odyssey
6	Color 1	Blue
7	Color 2	Red
8	Color 3	Silver
9	Color 4	Black
10	Color 5	Green

Truth-values are assigned to each of the color and style characteristics to create a feature vector that can be used to represent a buyer’s expression of their specific preference of automobile. The example assumes that people express their preference by buying cars from the hypothetical manufacturer. Their purchase preferences are represented by an ordered sequence of the ten truth-values in Table 2 formed into a feature vector on the basis of the characteristics of the purchased automobile.

The first hypothetical purchaser, “Gerald” bought a green Odyssey. The feature vector that represents his choice is illustrated with the one-dimensional array in Figure 1.

	F(1)	F(2)	F(3)	F(4)	F(5)	F(6)	F(7)	F(8)	F(9)	F(10)
Gerald	false	false	false	false	true	false	false	false	false	true

Figure 1. Feature vector for Gerald's auto preference

Each of our hypothetical buyers can be represented by an appropriate set of ordered truth-values specific to each individual’s selection of automobile. For the purposes of illustration six buyers are defined in Table 3.

Table 3. Auto Example Buyer Preferences

Buyer	Type of automobile purchased
Allen	Blue CRV
Marilyn	Silver Accord
Logan	Black Odyssey
Bonnie	Black Passport
Linda	Silver Civic

The probability model of buyer preferences, using the same feature set defined in Table 2 is illustrated in Table 4.

A probability model provides an estimate of the frequency of each type of automobile, given the known preferences of some sample set. There are many ways in which probabilistic induction can be used, however, probabilistic induction is not the focus of this research. The objective of this research is to focus on building models from a deterministic perspective.

Table 4. Probabilities of Buyer Preferences

	F(1)	F(2)	F(3)	F(4)	F(5)	F(6)	F(7)	F(8)	F(9)	F(10)
Gerald	0	0	0	0	1	0	0	0	0	1
Allen	0	0	0	1	0	1	0	0	0	0
Marilyn	0	1	0	0	0	0	0	1	0	0
Logan	0	0	0	0	1	0	0	0	1	0
Bonnie	0	0	1	0	0	0	0	0	1	0
Linda	1	0	0	0	0	0	0	1	0	0
Total	1	1	1	1	2	1	0	2	2	1
Probability	0.167	0.167	0.167	0.167	0.33	0.167	0.0	0.33	0.33	0.167

In the hypothetical group there is a 16.7 % probability that the manufacturer will sell a Passport (F3) to one of the buyers, and a 33% probability that the automobile will be black (F9). If we randomly select a buyer whose identity is not known and attempt to associate that buyer with a preference for a black automobile, while 1 out of 3 buyers in our example purchased a black automobile, it is not determinable whether our randomly selected buyer was one of the two who purchased a black car.

From the deterministic perspective, the only characteristic that every member of the group shares is that none preferred a red (F7) automobile. If our group is the entire distribution of automobiles, one can assert deterministically that the manufacturer didn’t sell any red vehicles. For the group represented, however, every style was selected by at least one person and four of the five possible colors were selected by at least one person.

The deterministic model focuses on contexts in which features are universally quantified. At the level of our group of six buyers, the truth-value associated with the selection of a red automobile is "*universally false*," because no one in our group selected a red automobile. For each of the other nine characteristics the truth-value used to represent the group must reflect that at least one true (selection = *true*) case exists. A single case demonstrates existential quantification. At the same time, at the group level we must represent each characteristic that is existentially *false* wherever there is a demonstrable false case. The selection of a blue auto, at the group level, is neither *universally true* or *universally false*, and so we shall treat that situation and any similar situation in which a feature may be demonstrated to be both *true* and *false*, depending upon which cases are inspected, as being "indeterminate."

In a collection where every feature of every case is known and assigned one of the alternative Boolean truth-values {*true*, *false*}, there are three possible truth-values that can be represented in an aggregate. These are illustrated in Table 5.

The meanings of *true* and *false* are universally quantified when applied to describing an individual or an aggregate. The *indeterminate* truth-value is meaningless in representing individuals and is only appropriate for an aggregate with two or more members. In the balance of this chapter and subsequent chapters for consistency, a *true* condition (individual or aggregate) will be represented by T or *true*, a *false* condition (individual or aggregate) will be represented by F or *false*, and aggregates that contain cases that are both *true* and *false* will be represented by I or *indeterminate*.

Table 5. Truth-Values Required to Describe Aggregates

Truth-value	Interpretation
True	True, universally true
False	False, universally false
Indeterminate	Sometimes true and sometimes false Existentially true and existentially false

The collection of six buyers, modeled using truth-values at both the individual and collection levels is illustrated in Table 6.

Table 6. Truth-Values for Individuals and Aggregates

	F(1)	F(2)	F(3)	F(4)	F(5)	F(6)	F(7)	F(8)	F(9)	F(10)
Gerald	F	F	F	F	T	F	F	F	F	T
Allen	F	F	F	T	F	T	F	F	F	F
Marilyn	F	T	F	F	F	F	F	T	F	F
Logan	F	F	F	F	T	F	F	F	T	F
Bonnie	F	F	T	F	F	F	F	F	T	F
Linda	T	F	F	F	F	F	F	T	F	F
Group	I	I	I	I	I	I	F	I	I	I

The three truth-values in Table 5 can only be applied to features that are known. To have a complete set of the possible types of truth-values we must also account for cases in which the truth-value of one or more features, for whatever reason, may not be known. This sort of situation can be demonstrated in the example, by adding another person to the auto buyer's group. Cecily will be added without specified preferences to illustrate handling missing information.

Accommodating the unknown is a strength of the probabilistic inductive model, provided that the behavior of unknown cases is statistically homogeneous with the behaviors of the cases used to formulate the model. The probabilistic model, given what we know from our group's experience and assuming that it is representative, might lead us to believe that there is an equal probability of Cecily selecting an Odyssey or a Civic and a slightly better probability that she might select a black or silver vehicle over one that is blue. Statistically we would not expect her to purchase a red car.

From a deterministic perspective, since we do not have any information about her selection, the truth-values we assign to each feature to represent the car Cecily selected, must reflect the fact that her choices are unknown. To accommodate missing information *unknown* is added to the deterministic truth-values in Table 6 to account for unknown or missing information.

Table 7. Deterministic Inductive Logic Truth-Values

Truth values	Meanings
T or *true*	True or universally true
F or *false*	False or universally false
I or *indeterminate*	Sometimes true and sometimes false Existentially true and Existentially false
U or *unknown*	Unknown

The relationship between known, unknown, true and false is represented in Table 8. The equivalent terms formed by negation are included in brackets. If we were to combine what we don't know about Cecily's preferences with what we do know about the balance of the group, from a deterministic inductive perspective, we obtain the generalization illustrated by the last row of Table 9. This describes the total group of seven buyers formed by combining Cecily's preferences with those of our six previous buyers.

Table 8. Relationships of Known and Unknown Truth-Values

And	Known	Unknown [~known]
True [~false]	True [~false]	Unknown [~known]
False [~true]	False [~true]	Unknown [~known]

Table 9. Representing Aggregates that Include Unknown Values

Feature	F(1)	F(2)	F(3)	F(4)	F(5)	F(6)	F(7)	F(8)	F(9)	F(10)
Original Group	I	I	I	I	I	I	F	I	I	I
Cecily	U	U	U	U	U	U	U	U	U	U
New Group	I	I	I	I	I	I	U	I	I	I

One alternative is to declare that whenever a single case's feature's value is *unknown* the collection's feature's value is *unknown*. This is clearly reasonable in the case of a collection that, without the presence of the unknown would have been either *universally true* or *universally false,* however, if one or more cases in the collection have the feature value as *true* and one or more cases has the feature value as *false,* there is no additional information that would result in any conclusion other than that the collection's feature value is mixed and therefore *indeterminate*.

We can segment a collection and define sub-contexts by, for example, segregating the buyers whose preferences we do not know from those that we have information about. The truth-value table that is used to support deterministic inductive logic must contain a minimum of four basic values from Table 7 above.[67]

The combination of two cases (individuals or aggregates) represented by the four truth-values results in an aggregate or group truth-value. The combinations of the four truth-values are defined in Table 10.

In the example, if we combine Logan and Bonnie into a group and formulate a deterministic descriptive feature vector for their group we obtain the results listed in the bottom line of Table 11.

---

[67] The base set of four values presumes that, at a minimum, any definable feature may be either present or absent in an object or concept being modeled. The representation of faceted structures with more than two mutually exclusive and collectively exhaustive characteristics is addressed in a later part of this paper.

Table 10. Truth-Value Combinations

Combination	T	F	I	U
True	T	I	I	U
False	I	F	I	U
Indeterminate	I	I	I	I
Unknown	U	U	I	U

Table 11. Group Feature Vector Defined by Buyers Logan and Bonnie

	F(1)	F(2)	F(3)	F(4)	F(5)	F(6)	F(7)	F(8)	F(9)	F(10)
Logan	F	F	F	F	T	F	F	F	T	F
Bonnie	F	F	T	F	F	F	F	F	T	F
Group	F	F	I	F	I	F	F	F	T	F

The group defined by Logan and Bonnie can be used to represent a hypothetical market segment. We will, for the sake of illustration, assume that the preference behaviors expressed by Logan and Bonnie collectively represent a "market segment" of buyers of automobiles. If the preferences of Logan and Bonnie are used to represent that market segment for our hypothetical automobile manufacturer, that manufacturer can never expect to sell a Civic, CRV or Accord or any automobile that is not black. If we conceptualized a salesperson standing at the end of the factory production line, segregating the cars that can be sold, based upon the Logan—Bonnie model, that salesperson would eliminate (not select) any car that was not black, however, that criteria (~black) would still leave automobiles, including black Civics, Accords and CRVs, that would not sell in our hypothetical market segment. The color black is clearly a feature that can be used to discriminate potentially sellable cars from ones that we would not expect to sell to folks whose preferences are characterized by the group defined by Logan—Bonnie.

To improve the quality of the end of production selection operation our salesperson could build a two-stage filter to select automobiles that are sellable in our hypothetical Logan—Bonnie market. We could represent the desirable subsets of cars, by first selecting cars that are black and then rejecting any cars that are not Passports or Odysseys. A deductive hierarchical classification of sellable cars is illustrated in Figure 2.

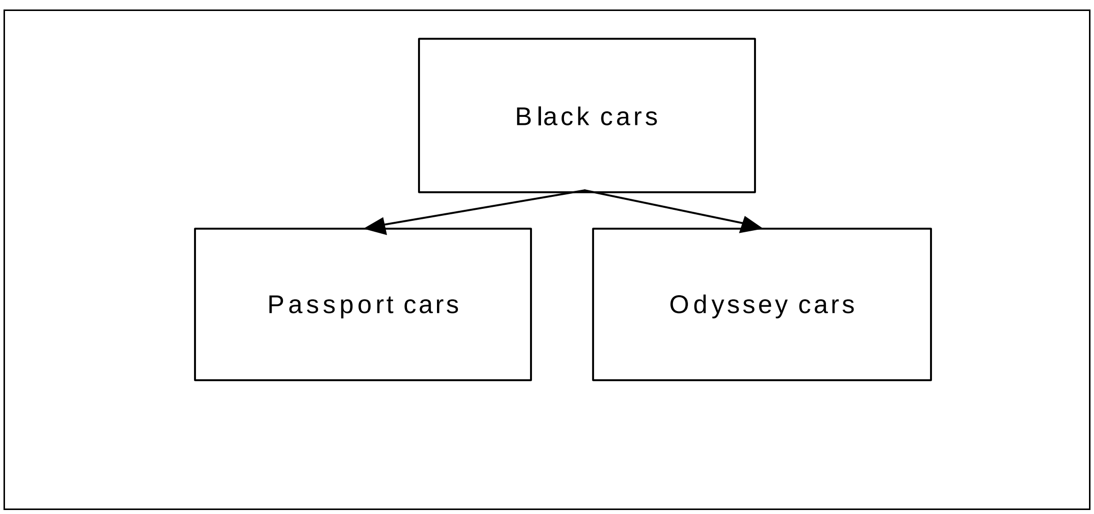

Figure 2. Filter model based on Logan and Bonnie preferences

If our hypothetical Logan--Bonnie group is conceived of as the universe, we can classify our purchasers into a hierarchy where the first differentiae is the color {black} of sellable cars and the second differentiae is the style {Passport, Odyssey} of sellable cars. Alternatively, we could accomplish the same objective by selecting style to be the first level differentiae and color to be the second. The change in sequence would result in the deductive classification in Figure 2 being reconfigured as illustrated in Figure 3.

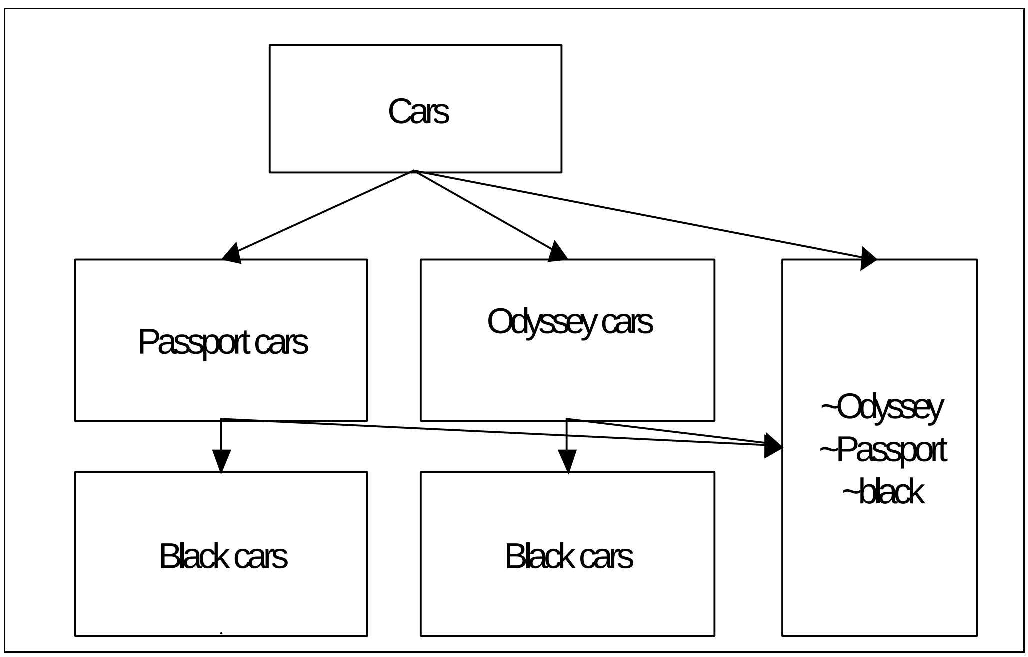

Figure 3. Classification based on Logan and Bonnie's preferences

If we were to account for every possible automobile in our example, as is done in probabilistic inductive models, we must account for five styles each of which might come in any one of the five colors. The combination results in twenty-five possible homogeneous types or categories of automobiles. That general model, while it accommodates any possible combination of features that might be represented by any single buyer or group of buyers is far larger and would require more computational effort to track than the deterministic contextual model defined by combining the preferences of Logan and Bonnie.

We can only construct a probabilistic inductive model to describe six of our seven buyers. We cannot represent Cecily's unknown preferences in the quantitative model. In Table 12, each buyer's preference is identified. The bottom row represents the probabilities by color selections and the rightmost column the probabilities by style. The buyers are identified in the cells associated with their appropriate style and color preferences.

Table 12. Buyer Preference Probability Model

	Blue	Red	Silver	Black	Green	Style
Civic			1 Linda			0.167
Accord			1 Marilyn			0.167
Passport				1 Bonnie		0.167
CRV	1 Allen					0.167
Odyssey				1 Logan	1 Gerald	0.33
Color	0.167	0.0	0.33	0.33	0.167	1.0

The quantitative model allows us to trace every possible combination of color and style and predict, if our sample is representative, the preferences of buyers like Cecily whose preferences are not known. However, if we are only interested in a sub-context of the possible and are willing to give up the universality of the quantitative inductive model, we can represent a limited context of interest by reducing the size of the model. This results in a more computationally efficient representation that ignores non-essential or uninteresting information.

In the case of the market segment characterized by Logan and Bonnie, the model can be reduced to an array with two dimensions, with three “styles” in one dimension and two “colors” in the other.[68] The model illustrated in Table 13 requires only six cells to account for all the relevant output of the manufacturer. Two cells account for purchaser preferences demonstrated by Logan and Bonnie and four cells account for all other automobiles produced by our hypothetical manufacturer.

Table 13. Reduced Logan-Bonnie Model

	Black	not(Black)
Passport	1 Bonnie	
Odyssey	1 Logan	
not(Passport or Odyssey		

[68] We have changed the meaning of style and color in this example. We are, in this example, reconstructing the representation for the purpose of illustration. The representation of the reduced model is intended only to reflect a sufficient number of features to discriminate between Logan and Bonnie and between the market segment defined by combining Logan and Bonnie from any other possible market segment.

We could just as easily have selected a group that was characterized by Allen and Gerald. That market segment's deterministic representation is illustrated in Table 14.

Table 14. Market Segment Preferences Defined By Gerald And Allen

	F(1)	F(2)	F(3)	F(4)	F(5)	F(6)	F(7)	F(8)	F(9)	F(10)
Gerald	F	F	F	F	T	F	F	F	F	T
Allen	F	F	F	T	F	T	F	F	F	F
Group	F	F	F	I	I	I	F	F	F	I

The Gerald-Allen deductive classification structure must account for two of the five possible styles and two of the five possible colors. The Gerald-Allen model illustrated in Figure 4 is a slightly more complex deductive tree than was illustrated in Figures 2 and 3 to segregate the Logan-Bonnie market segment of automobiles.

If we defined sellable cars in a market segment as "relevant objects," we could consider the ones that we wouldn't sell in a particular market segment as "not-relevant objects." If we use the preferences of Gerald and Allen to select relevant objects (sellable cars), we can segregate or "filter out" all the non-relevant objects by excluding all Accords, Passports, and Civics, and all red, silver or black cars.

The example is intended to illustrate how inclusionary and exclusionary criteria can be obtained from group characterizations formed using the deterministic inductive combination approach. An *indeterminate* truth-value cannot be used as exclusionary criteria because sometimes objects that express the features that are *indeterminate* at the group level are “relevant” for an individual or subset of the group. *True* is a universal inclusionary criteria for “relevant” objects, while *false* is a universal exclusionary criteria for “relevant” objects.

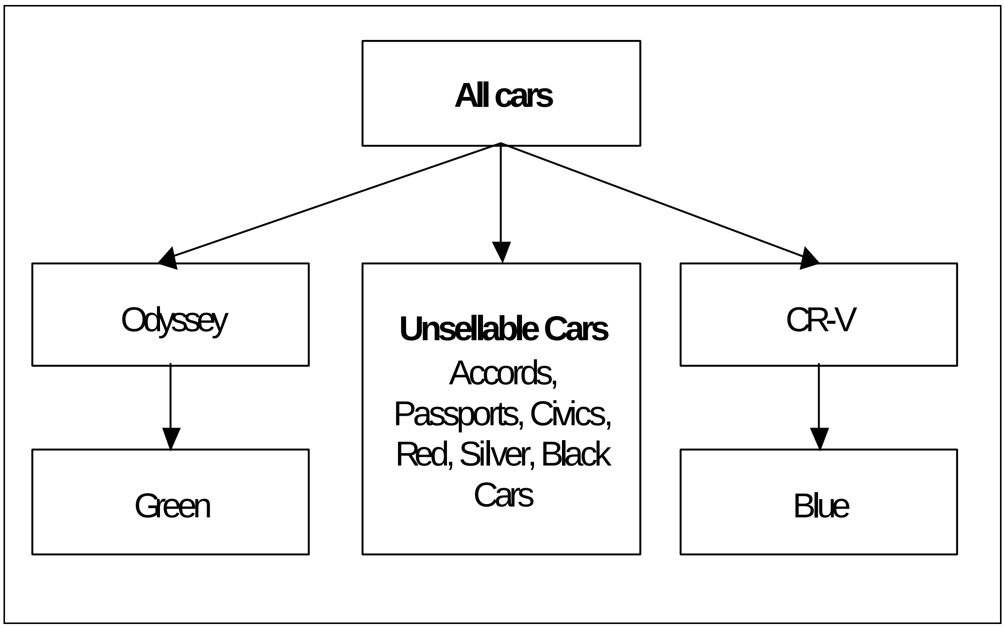

Figure 4. Classification based on Gerald and Allen’s preferences

Inductive classification

Each of the classifications above has been described as a deductive classification because each begins with a universe and reduces that universe using successive division. If instead of trying to account for all possible cars one were to build an inductive classification on the basis of the specific automobiles and groups of automobiles of interest, the classification structure would change. First, the top of the classification would not represent a universe, but rather some number of possible generalizations that can be formed by combining different individual cases and groups formed by combining cases.

Figure 5 illustrates an inductive classification formed on the basis of the purchase preferences of Logan and Bonnie. The possible categories or generalizations are constructed from the bottom-up instead of from the top-down as was illustrated in Figure 3. The top row of categories in Figure 5 represent generalizations that are homogeneous that can be inferred from the characteristics of the two automobiles that were used to define the market segment.

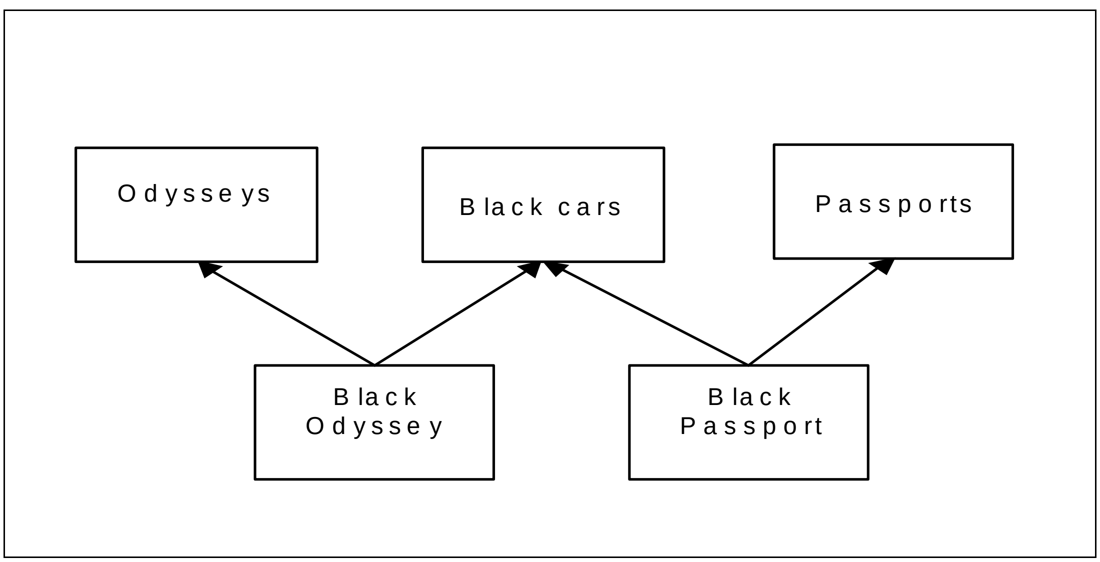

Figure 5. Logan-Bonnie market segment inductive classification

Enumeration features

The identity of cases that are associated can be used to define characteristics of a group in terms of members of that group. The identity of a buyer may be a useful basis for grouping in addition to object characteristics like style and color. Buyers' identities can be added to the vocabulary by listing them and assigning each a feature number and a truth-value. Buyer identity, for example, may become useful in incorporating buyer characteristics into the model such as income, national origin, age, etc, to understand the characteristics of the market segment not only in terms of the objects but also in terms of associated characteristics. Enumerating buyers is illustrated in Table 15.

Table 15. Adding Features by Enumeration

Case Identification	F(11)	F(12)	F(13)	F(14)	F(15)	F(16)	F(17)
Gerald	T	F	F	F	F	F	F
Allen	F	T	F	F	F	F	F
Marilyn	F	F	T	F	F	F	F
Logan	F	F	F	T	F	F	F
Bonnie	F	F	F	F	T	F	F
Linda	F	F	F	F	F	T	F
Cecily	F	F	F	F	F	F	T

An eighteenth feature might be defined to record the manufacturer of automobiles if the intent were to extend the model to additional manufacturers or simply to make manufacturer explicit. To be meaningful and comparable, every feature-vector must be sequenced in the same order so that the semantics of features remain consistent. The model becomes meaningless if, for example, sometimes feature number 15 is used to represent the buyer "Bonnie" and sometimes it is used to represent the buyer "Gerald."

Every inductive model requires a mechanism to record the meaning of every feature or cell in every feature-vector. That mechanism can be conceptualized as a registry, where features are registered and (1) associated with a numeric cell in the feature-vector and (2) given a definition or scope note that explains how one assigns truth-values to that feature.

Generalized combine operator for multi-faceted features

If instead of defining the seven buyers as seven individual features, we were to define a characteristic "buyer" and give it seven truth-values, we would need a capability to combine faceted structures that are not binary. Every faceted model, like the binary model requires that each one of *n* features be selected from a set of features that are mutually exclusive and collectively exhaustive. Examples of faceted truth tables with different numbers of facets is illustrated in Table 16.

The first example in Table 16 has two facets (N=2), and is an example of a Boolean set of mutually exclusive and collectively exhaustive alternatives {*true*, *false*}. The second example might be used in a five choice multiple-choice examination, where each value has a different meaning. To accommodate missing information, the *unknown* truth-value serves the same role and purpose in a multiple-choice situation as it did in a Boolean or binary choice situation. The combine operator, to be consistent in both binary and multiple facet models, must have an *indeterminate* truth-value to represent a feature that is heterogeneous in an aggregate.

Table 16. Multi-Faceted Features

Feature vector sequence number	Set of possible assignable truth-values	Rules for assigning truth-values
N=2	True [~false] False [~true]	True—when feature is present False—when feature is absent
N=5	TV(mc1) TV(mc2) TV(mc3) TV(mc4) TV(mc5)	For Question_____ TV(mc1)—when the oval associated with multiple choice 1 is marked TV(mc2)—when the oval associated with multiple choice 2 is marked TV(mc3)--when the oval associated with multiple choice 3 is marked TV(mc4)—when the oval associated with multiple choice 4 is marked TV(mc5)—when the oval associated with multiple choice 5 is marked
N + 1	Indeterminate	Existentially any one of N in a group containing more than one case
N + 2	Unknown	Unknown facet

Table 17 illustrates the deterministic inductive combination operator that would be used to model a four-facet feature such as the chemicals present in DNA base pairs {A, G, C, T}.[69]

---

[69] Each chromosome is made up of a molecule of DNA in the shape of a double helix which is composed of four chemical bases represented by the letters

Table 17. Combine for the DNA Four-Facet Base-Pair Model

Combine	A	G	C	T	Indeterminate	Unknown
A	A	I	I	I	I	U
G	I	G	I	I	I	U
C	I	I	C	I	I	U
T	I	I	I	T	I	U
Indeterminate [I]	I	I	I	I	I	I
Unknown [U]	U	U	U	U	I	U

The combine operator can be generalized for any number of facets by adding to the list of mutually exclusive and collectively exhaustive facets the *indeterminate* and *unknown* facets required by aggregates formed with the combine operator of deterministic inductive logic. The faceted form of the combine operator is illustrated using a human readable pseudocode in Figure 6.

A (adenine), G (guanine), C (cytosine) and T (thymine). The arrangement or sequences of the letters determines the cell's genetic code.

```
Function COMBINE (argument case-list) Returns feature-vector

Initialize an aggregate feature vector as the equivalent of the first case that
is to be aggregated

For each additional case in the case-list that is to be aggregated
For each feature in each additional case's feature vector
    If the truth-value in the cell of the aggregate feature-vector
        IS IDENTICAL TO
            the truth-value in the respective cell of the additional case's
                feature-vector
        THEN
            Retain the value in the aggregate feature-vector
        ELSE
            If the truth-value in the cell of the aggregate feature-vector
                IS EQUIVALENT TO Unknown
            THEN
                Set the truth-value in the aggregate feature-vector to Unknown
            ELSE
                If the truth-value in the respective cell of the additional
                        case's feature-vector IS EQUIVALENT TO Unknown
                    AND
                        the truth-value in the cell of the aggregate feature-vector
                        IS NOT EQUIVALENT TO Indeterminate
                THEN
                    Set the truth-value in the aggregate feature-vector to Unknown
                ELSE
                    Set the truth-value in the aggregate feature-vector to
                        Indeterminate
                ENDIF non-identical values
            ENDIF identical values
        NEXT feature
NEXT case

END FUNCTION
```

Figure 6. Combine function pseudocode (*n-facet* version)

The pseudocode illustrated in Figure 6 can be translated into a language like FORTRAN or C for implementation and execution in a computer. All array data structures have the potential to be parallelized to improve execution speed when the appropriate hardware is available for operations. Since deterministic inductive logic is based upon array data structures and algorithms that manipulate arrays, deterministic inductive logic can potentially be implemented and executed in synchronous and asynchronous parallel systems.

Comparing and contrasting feature vectors

Deterministic inductive logic uses individual cases to create specifications of sets of cases by combination to represent aggregates. The aggregates are assumed to be based upon some intentional selection of cases for combination. The combine operator facilitates the creation of aggregate feature vector specifications to represent new concepts by example.

To be analytically functional, a method for comparing and contrasting categories and cases to assess how the features of individual cases and case aggregates are similar or different is required. Two deterministic inductive logic functions are defined to facilitate vector comparisons:

1. Compare—identify identical features in two feature-vectors
2. Contrast—identify non-identical features in feature-vectors

The compare operator returns *true* when identical values are compared and *false* when the values are different. The contrast operator returns *false* when identical values are compared and *true* when the values are different. The compare operator is illustrated for the four-value form of deterministic inductive logic in Table 18 and the contrast operator in Table 19.

Table 18. Compare Operator

Compare	T	F	I	U
True	T	F	F	F
False	F	T	F	F
Indeterminate	F	F	T	F
Unknown	F	F	F	T

Table 19. Contrast Operator

Contrast	T	F	I	U
True	F	T	T	T
False	T	F	T	T
Indeterminate	T	T	F	T
Unknown	T	T	T	F

For an *n-facet* model, the compare and contrast functions must be implemented with a computer algorithm to permit comparing and contrasting lists of facet values. The compare function is illustrated using a human readable pseudocode in Figure 7 and the contrast function is similarly illustrated in Figure 8.

Function COMPARE (***argument*** *case-list*) **Returns** *resultant-comparison-feature-vector*

Initialize a *base-comparison-feature-vector* as the equivalent of the first case that is to be compared

Define *no-further-comparison-required* as a special facet to be used whenever two respective cells of any two compared feature-vectors are not identical

Initialize a *resultant-comparison-feature-vector* with all cells to the truth-value *false*

```
For each additional case in the case-list that is to be compared
    For each feature in each additional case's feature vector
        If the truth-value in the cell of the base-comparison feature-vector
            IS NOT IDENTICAL TO no-further-comparison-required
            AND
                the truth-value in the cell of the base-comparison feature-vector
                    IS IDENTICAL TO
                    the truth-value in the respective cell of the additional case's
                        feature-vector
        THEN
            Set the truth-value of resultant-comparison-feature-vector to the
                Truth-value true
        ELSE
            Set the truth-value of resultant-comparison-feature-vector
                to the truth-value false
            Set the truth-value in the respective cell of the
                base-comparison feature-vector
                    to the truth-value no-further-comparison-required
        ENDIF comparison
    NEXT feature
NEXT case
```

END FUNCTION

Figure 7. Compare function pseudocode

Function CONTRAST (***argument*** *case-list*) **Returns** *resultant-contrast-feature-vector*

Initialize a *base-contrast-feature-vector* as the equivalent of the first case that is to be contrasted

Define *no-further-contrasting-required* as a special facet to be used whenever two respective cells of any two compared feature-vectors are identical

Initialize a *resultant-contrast-feature-vector* with all cells to the truth-value *true*

```
For each additional case in the case-list that is to be contrasted
    For each feature in each additional case's feature-vector
        If the truth-value in the cell of the base-contrast-feature-vector
                IS IDENTICAL TO no-further-contrast-required
            AND
                the truth-value in the cell of the base-comparison feature-vector
                    IS NOT IDENTICAL TO
                        the truth-value in the respective cell of the additional case's
                            feature-vector
        THEN
            Set the truth-value of resultant-comparison-feature-vector to the
                Truth-value false
        ELSE
            Set the truth-value of resultant-comparison-feature-vector
                to the truth-value true
            Set the truth-value in the respective cell of the
                base-comparison feature-vector
                to the truth-value no-further-contrast-required
        ENDIF contrast
    NEXT feature
NEXT case
```

END FUNCTION

Figure 8. Contrast function pseudocode

Negation

Deterministic inductive logic supports negation for binary–based faceted information structures but not for non-binary multi-faceted information structures. In a binary-faceted model, to be consistent with deductive logic it is desirable for *true* = ~*false*. In the case of *unknown,* while it might be reasonable for ~*unknown* to be any facet other than *unknown*, to preserve consistency *unknown* = ~*indeterminate* and ~*indeterminate* = *unknown.*

Table 20. Negation for Binary Features

Facet	Negation
true ; always true	*false*
false ; always false	*true*
indeterminate	*unknown*
unknown	*indeterminate*

In a multi-faceted information structure, negation is supported for the *indeterminate* and the *unknown* facets associated with an *n-facet* information structure as illustrated in Table 21.

Table 21. Negation for *Indeterminate* and *Unknown* Facets in *n-Facet* Features

TV(n+1); *indeterminate*	TV(n+2); *unknown*
TV(n+2); *unknown*	TV(n+1); *indeterminate*

There are potentially some situations in which one can define, within a specific faceted structure, rules for selecting the negation of a facet, however, these rules are specific to each faceted structure and the semantics of its implementation. For example, in a set of five possible outcomes, such as in a multiple choice test, the concept of "not outcome 5" is not any one facet, but could be any other member of the set except number five. In situations in which there are no specific rules for understanding negation in a non-binary faceted information structure, an alternative that supports negation is to represent the *n-facets* as *n* separate features. In the buyer example, we can assign the seven buyers to seven facets of a single feature or to seven features where the set of features is associated with an enumerated list of possible alternative values.

The negation of compare is contrast and the negation of contrast is compare. Thus ~COMPARE(feature vector(1), feature vector(2)) = CONTRAST(feature vector(1), feature vector(2)) and ~CONTRAST(feature vector(1), feature vector(2)) = COMPARE(feature vector(1), feature vector(2)). These truth-values are illustrated in Table 22.

Table 22. Negations of Truth-Values in Compare and Contrast Operators

Function	Truth-value	Truth-value negation
Compare	*true*	*false*
Compare	*false*	*true*
Contrast	*true*	*false*
Contrast	*false*	*true*

Basic feature-vector types

There are four basic vector types that are used in formulating statements in deterministic inductive logic. These are:

1. *Case-vector*
2. *Generalization-vector*
3. *Identity-vector*
4. *Difference-vector*

The *case-vector* has been described. It is the mechanism used to record the characteristics of an individual object or case.

The *generalization-vector* is defined by combination. The generalization-vector can be used, for example, to test two vectors for hierarchical subsumption, where one of the vectors subsumes the other on the basis of the features each expresses. A hierarchical generalization can be formed to subsume any two feature-vectors by combining those two feature vectors. For example, a generalization that subsumes both *feature-vector-XX* and *feature-vector-XY* can be derived by combination.

Generalization-XX+XY = COMBINE(*feature-vector-XX*, *feature-vector-XY*)

Figure 9. Generalization formula

The role of an *identity-vector* is to provide a basis of comparison for two vectors. An *identity-vector* is derived from the comparison of any feature vector with itself.

Identity-vector = COMPARE(*feature-vector-XX*, *feature-vector-XX*)

Figure 10. Formula for creating identity vectors

The negation of an *Identity-vector* is a *difference-vector*. A *difference-vector* is derived from the contrast of any feature vector with itself.

Difference-vector = CONTRAST(*feature-vector-XX*, *feature-vector-XX*)

Figure 11. Formula for creating difference vectors

Discriminants

A *discriminant* is a homogeneous (universally quantified) characteristic in any aggregate. The term *discriminant* is not applied to individual *case-vectors*, but is reserved for *generalization-vectors*. In a binary logic, discriminants can be either inclusionary (*true)* or exclusionary (*false)*. A *discriminant-vector* is created to identify those features that are either inclusionary or exclusionary discriminants. A *discriminant-vector* is a derived vector type.

We originally introduced the concept of discriminants in the Logan-Bonnie group of automobiles where the one feature common to all the members of the group was the color black. The color black (feature F(9)) is the sole inclusionary discriminant defined by the combination of Logan and Bonnie.

Table 23. Logan and Bonnie Combination Vector

	F(1)	F(2)	F(3)	F(4)	F(5)	F(6)	F(7)	F(8)	F(9)	F(10)
Logan	F	F	F	F	T	F	F	F	T	F
Bonnie	F	F	T	F	F	F	F	F	T	F
Group	F	F	I	F	I	F	F	F	T	F

Features {1,2,4,6,7,8,10) in Table 23 are exclusionary discriminants. Features 3 and 5 in Table 23 are not discriminants, as their value is *indeterminate* with respect to the group defined by Logan and Bonnie.

Figures 12 and 13 contain the formulas used to construct the two types of *discriminant-vector* used in deterministic inductive logic.

Inclusionary-discriminant-vector = COMPARE(COMBINE*(case-list)*, *identity vector*)

Figure 12. Formula for creating an inclusionary discriminate vector

Exclusionary-discriminant-vector = COMPARE(COMBINE*(case-list)*, *difference-vector*)

Figure 13. Formula for creating an exclusionary discriminate vector

In each case the discriminants in the resulting vector will have the value *true* because of the formulation of the comparison. Table 24 illustrates creating an *Inclusionary-discriminant-vector* and Table 25 illustrates creating an *Exclusionary-discriminant-vector*.

Table 24. COMPARE(COMBINE*(case-list)*, *identify vector*) Truth Table

Group	F	F	I	F	I	F	F	F	T	F
Identity	T	T	T	T	T	T	T	T	T	T
Inclusion	F	F	F	F	F	F	F	F	T	F

Table 25. COMPARE(COMBINE*(case-list)*, *difference-vector*) Truth Table

Group	F	F	I	F	I	F	F	F	T	F
Difference	F	F	F	F	F	F	F	F	F	F
Exclusion	T	T	F	T	F	T	T	T	F	T

Testing for hierarchical subsumption

Two conditions can be derived from the way the combine function is used to define an aggregate or class: (1) the constituents of the case list used to define a class are members of the combined class; and (2) each individual member of the case list that was combined to form a class is subsumed by that class. Figure 14 illustrates the test for hierarchical subsumption.

case(X) is subsumed by COMBINE (*case-list*)   **IFF** COMBINE (*case-list*, *case(X)*) = COMBINE (*case-list*)

Figure 14. Test for hierarchical subsumption

While we can certainly track members of enumerated lists to assess whether objects are members of sets, the objective is to make such assessments on the basis of the characteristics that objects present. We are interested in the features that characterize an aggregate and in ways we can use what we know about the characteristics of objects and aggregates to determine whether an object that is not a member of a known enumerated set should belong to that set on the basis of the characteristics it expresses.

For example, we could define an aggregate in our automobile preference example by combining the preferences expressed by Gerald (green Odyssey) and Bonnie (black Passport) (see Table 26) and then test to see if Logan is conceptually a member of the market segment defined by Gerald-Bonnie (see Table 27).

As illustrated in Table 28, the comparison of the two aggregates {GB, GBL} resulted in the identity-vector, which is interpreted as meaning that the aggregates defined by the two different case-lists are conceptually equivalent. Another way to describe these two concepts {GB, GBL) is as synonyms.

Table 26. Gerald-Bonnie Combination

COMBINE	F(1)	F(2)	F(3)	F(4)	F(5)	F(6)	F(7)	F(8)	F(9)	F(10)
Gerald	F	F	F	F	T	F	F	F	F	T
Bonnie	F	F	T	F	F	F	F	F	T	F
GB	F	F	I	F	I	F	F	F	I	I

Table 27. Gerald-Bonnie-Logan Combination

COMBINE	F(1)	F(2)	F(3)	F(4)	F(5)	F(6)	F(7)	F(8)	F(9)	F(10)
GB	F	F	I	F	I	F	F	F	I	I
Logan	F	F	F	F	T	F	F	F	T	F
GBL	F	F	I	F	I	F	F	F	I	I

Table 28. Test for Inclusion of Logan in the Gerald-Bonnie Segment

COMPARE	F(1)	F(2)	F(3)	F(4)	F(5)	F(6)	F(7)	F(8)	F(9)	F(10)
GB	F	F	I	F	I	F	F	F	I	I
GBL	F	F	I	F	I	F	F	F	I	I
COMPARE(GB, GBL)	T	T	T	T	T	T	T	T	T	T

The subsumption tests can be generalized from the example into two statements formulated in terms of the combine and compare functions and the two tests can be used for testing the relationship between any two sets of objects in terms of their characteristics as illustrated in Figure 15.

IF COMBINE (*case-list(1)*, *case-list(2)*) = COMBINE (*case-list(1)*) **THEN**

COMBINE (*case-list(2))* **is subsumed by** COMBINE (*case-list(1)*)

IF COMBINE (*case-list(2)*, *case-list(1)*) = COMBINE (*case-list(2)*) **THEN**

COMBINE (*case-list(1))* **is subsumed by** COMBINE (*case-list(2)*)

Figure 15. Set relationship subsumption tests

The deterministic inductive logic capability to assess hierarchical subsumption results in the capability to test concepts for inclusion within other concepts. In our example, we can extend the concept list by adding categories {GL, LB, GB} and conclude GBL is subsumed by GB by testing for hierarchical subsumption. The diagram in Figure 16 illustrates these relationships.

In the example illustrated in Figure 16, the combination of Logan and Bonnie resulted in a discriminant (black) that is a basic color concept from the initial set of ten style and color concepts. Both of the categories Logan-Bonnie and Gerald-Logan can be represented by inclusionary discriminants from the original set of ten terms. The combination of Gerald and Bonnie, however, represents a new concept that cannot be defined by inclusionary discriminants. To define the category that results by combining Gerald and Bonnie we use exclusionary discriminants and exclude all Accord, Civic, and CRV style cars and all Blue, Red and Silver cars.

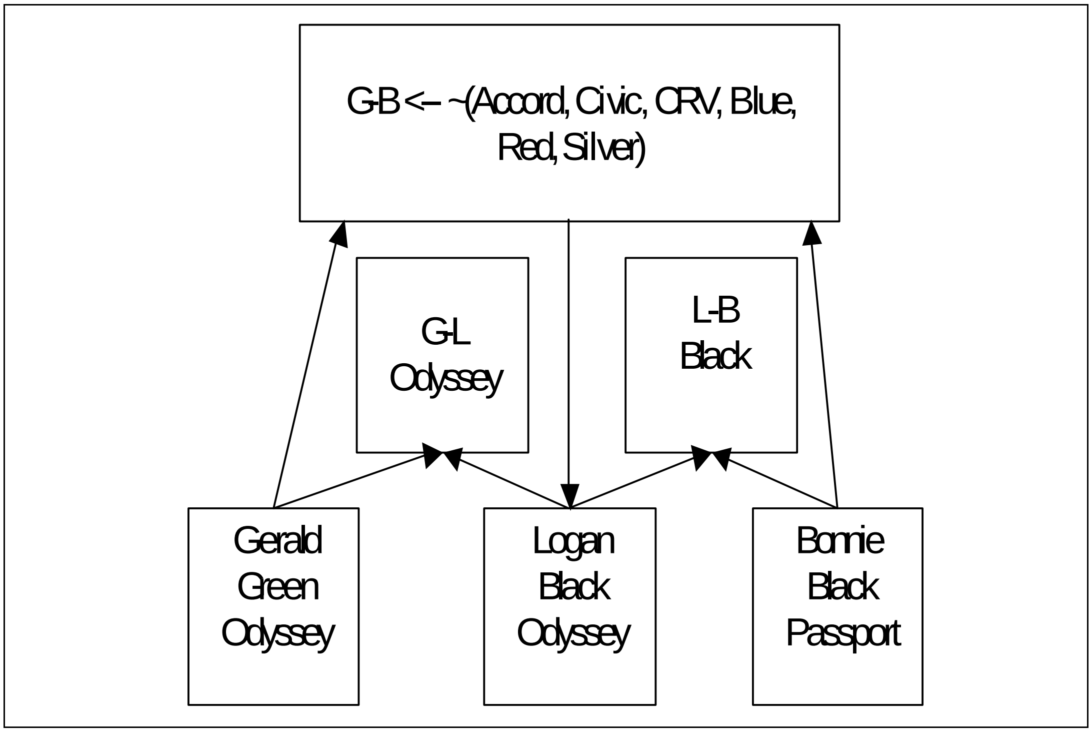

Figure 16. Class relationships formed inductively from individuals

The deductive logic forms of the categories in Figure 16 are:

- Logan (Black and Odyssey),
- Bonnie: (Black and Passport),
- Gerald: (Green and Odyssey),
- Gerald—Logan: (Black or Green) and (Odyssey)
- Logan—Bonnie: Black and (Odyssey or Passport)

- Gerald—Bonnie: (Green or Black) and (Odyssey or Passport)

The example illustrates how a set of observations can be combined in various ways to produce an inductive classification and how the new terms defined inductively can be represented in deductive logic using the original ten-term vocabulary.

Testing feature vectors for equivalence

To evaluate the equivalence between two vectors we need a measure of similarity or a measure of difference that can provide a basis for assessing whether two vectors are the same. If two vectors are equivalent, every feature in the two vectors must be identical. We can assess vector equivalence by comparing two vectors and then calculating their *degree of similarity* (DOS) or *degree of difference* (DOD). If any two vectors are 100% similar or 0% different, they are identical.

DOS – degree of similarity is determined by dividing the number of *true* truth-values in a vector created by a compare function by the total number of truth-values in that vector.

DOD – degree of difference is determining by dividing the number of *false* truth-values in a vector created by a contrast function by the total number of truth-values in that vector.

For two vectors (represented by *a* and *b*) being compared:

Equality: DOD(COMPARE(a, b)) = 0.00

DOS(COMPARE(a, b)) = 1.00

Inequality: DOD(COMPARE(a, b)) > 0.00

DOS(COMPARE(a, b)) < 1.00

The two functions that return values for degrees of similarity DOS and degrees of difference DOD are illustrated using pseudocode in Figures 17 and 18.

```
Function DOS (arguments vector(1), vector(2)) Returns dos

Initialize total-cells = 0
Initialize total-tv1 = 0
For each feature in vector(1)
    total-cells = total-cells + 1
    If the truth-value in the cell of vector(1)
        IS IDENTICAL TO
            the truth-value in the cell of vector(2)
    THEN
        total-tv1 = total-tv1 + 1
    ENDIF identical values test
NEXT feature

RETURN dos = total-tv1 / total-cells

END FUNCTION
```

Figure 17. Degree of similarity function

```
Function DOD (arguments vector(1), vector(2)) Returns dod

Initialize total-cells = 0
Initialize total-tv2 = 0
For each feature in vector(1)
    total-cells = total-cells + 1
    If the truth-value in the cell of vector(1)
        IS IDENTICAL TO
            the truth-value in the cell of vector(2)
    THEN
        total-tv1 = total-tv1 + 1
    ELSE
        total-tv2 = total-tv2 + 1
    ENDIF identical values test
NEXT feature

RETURN dod = total-tv2 / total-cells

END FUNCTION
```

Figure 18. Degree of difference function

Chapter IV – Formulating a vocabulary with deterministic inductive logic

This chapter addresses the formulation of feature vectors from both structured and unstructured information.

Vocabulary based systems

A vocabulary-based system is one in which a lexicon is developed to identify the concepts that are used in conjunction with search, retrieval and analysis. The terms in the vocabulary are the basis for formulating statements that express search, retrieval and analytical objectives. Vocabularies can be used in conjunction with both structured and unstructured information. The vocabulary associated with structured information might include a database schema, an information element name, a variable, or any unit of information that is referred to by or with computer software. In unstructured information, the most common use of a vocabulary is for subject, topic or other conceptual indexing.

Structured information systems are frequently associated with database management systems that use database schemas to define the system lexicon and the relationships between and among information elements. Unit record systems are an example of structured information systems in which fields are concatenated into elements and elements are concatenated into records and records are concatenated into files and files are collected into file systems. A payroll system, for example, might have a number of files, each containing multiple records, each record containing multiple information elements many of which might contain multiple fields that have a meaning that must be interpreted in context. For example, a date is an information element. In many cases dates are structured from fields that record respectively month, day and year in the context of a calendar (i.e., a Gregorian calendar) that was generally not explicitly specified. Information elements that may be associated with more than one time point may contain multiple dates in fields that are associated with the respective date meanings.

Unstructured information is frequently constructed from text corpora. Such systems are commonly referred to as information retrieval systems to distinguish them from structured information in database management systems. An information retrieval system generally manages what are termed *collections* in which the unit of storage in a collection is generally either a *document* or an *object*. In text collections the unit of storage is generally described as a document. Documents ordinarily contain information encoded in a language used by humans to communicate concepts in writing. The unit of storage in a non-text context, such as a multi-media collection, is commonly referred to as an *object*. A multi-media object is typically a unit of storage that can be performed for a user. It might contain audio and visual elements including pictures, graphics, text, music, speech, etc. The term *object* is also used to refer to computer software components and is a general term that can be used to describe any unit of content in an information system.

Many deployed information systems incorporate both structured and unstructured information. Structured information tends to be referred to with terms that might be variable names, object names, record identifiers, file identifiers, etc. An object's unstructured component (i.e., graphic, audio track, video track, etc.) that is to be retrieved and displayed or performed for a user must have an object name or identifier with which to reference and retrieve it. Since objects may contain objects and an object may require multiple components (i.e., audio track synchronized with a slide presentation) to be complete, every object identifier is a name used to refer to a container the holds either structured or unstructured or both structured and unstructured information.

The mechanisms used to retrieve object containers are dependent upon some form of description that a retrieval system can use to select objects. These descriptive mechanisms can be composed from the characteristics explicitly expressed by objects or by descriptive surrogates assigned by indexing systems. Descriptive surrogates may be implemented using quantitative methods, terms assigned by human indexers, or a combination of human and automated indexing sources. Automated indexing using language based terms and phrases exploits characteristics of objects that are cognitively accessible to users and are explicitly expressed in objects. Descriptive surrogates produced by human indexers are composed from indexing languages (controlled vocabularies) that are constructed to represent a particular configuration of concepts that are considered useful for managing information. While it might be possible to assign controlled vocabulary terms using a statistical classifier, a *controlled vocabulary* is commonly associated with descriptive surrogates assigned by human indexers. Irrespective of the methods of assignment, a controlled vocabulary based system is one that employs descriptive surrogates that are assigned by implication (e.g., they are not expressed explicitly in objects) to reflect conceptually accessible categories with which one may classify objects.

A term in a controlled vocabulary may also be used in full-text indexing. For example, a term like "records management" might be used both in full-text indexing and to describe a conceptual category assigned by inference. A record that contains the phrase "records management" might not be assigned the descriptive surrogate *records management*, if the indexing mechanism, such as the individual assigning controlled vocabulary terms, concluded that its assignment was not appropriate given the conditions for its use. For example, an article that only peripherally addresses records management that is primarily focused on the specifications and use of titanium bolts for securing cabinets in a records management facility, might be accessible when the phrase "records management" was used for full-text retrieval and yet not be retrieved by the descriptive surrogate *records management*.

Where an information retrieval system contains searchable characteristics derived from the characteristics expressed explicitly in objects and ones implicitly assigned by human indexers, such as Dialog, the system may provide the capability for users to search both types of characteristics. In Dialog, a searcher can formulate a search request in terms of both an index language and formulations of terms and the relative proximity of those terms in the object's full-text. Additionally, structured information elements like date of publication, author, publisher, language, etc., can be used in some systems to refine a query.

The specification of the subset of a collection of interest is termed a *query*. The formulation of a query is dependent upon how a retrieval system is designed to access information. If the system is accessible using a vocabulary, then a *query* would ordinarily be formulated by connecting terms in a vocabulary with Boolean and proximity operators. In systems that are indexed using quantitative methods, vocabulary based queries may be used as starting points to map a user's query expression into some entry point in the quantitative model or might be used to calculate objects association with the query.[70]

Irrespective of how a query is formulated, the query is the specification used to retrieve a set of objects that is a subset of all of the possible objects in a collection. Retrieval systems use the query as the basis for including and/or excluding objects in the results set. Systems, such as Dialog, allow users to combine two or more result sets. The combined set can be expressed by a

---

[70] In a ranked retrieval system, objects are ordered by a metric that is used to predict the probability that an object is relevant given a query. A similarity metric such as a cosine coefficient is calculated from the query's and objects' vector representations to reflect the relative relevance or rank of objects.

disjunction (Boolean “or”) of whatever queries were used to retrieve the individual sets that were combined.

Hierarchy

Controlled vocabulary systems can be differentiated in terms of how the concepts are selected. Document oriented indexing languages are constructed of terms to represent the concepts expressed in objects in a collection, where the focus of the effort is to produce a schema of concepts that reflects the structure of the information contained in information objects. An alternative approach that focuses on how users conceptualize their information needs seeks to construct a schema (user oriented indexing language) that can be used to express and classify the information needs of users that motivate information search processes. In practice, lexicographers will generally seek a balance between their understanding of the needs users will present to the system and the concepts presented by authors in documents in the collection.

A controlled vocabulary system’s operation is based upon matching terms and term relationships. Matching requires a consistency between the terms used in queries and those included in descriptive surrogates. The terms that are permitted to be used in queries and object surrogates are known as *preferred terms*. Preferred terms make up the indexing language that is used to discriminate relevant objects. A preferred term is included in an object surrogate to indicate that object’s applicability or relevance with respect to a user need expressed using the same term. For example, a query requesting objects relevant to *quantum mechanics* would search for objects whose surrogates contained the entry quantum mechanics. Similarly, a query that requested *physics and not quantum mechanics* would search for objects whose surrogates indicated they are applicable to the subject *physics* but would include only those objects whose surrogates do not contain the entry *quantum mechanics*.

Controlled vocabulary systems restrict the terms that can be used in both queries and object surrogates. Users that are not intimately familiar with a system's vocabulary need to translate their information needs into the system's vocabulary. A useful component of a controlled vocabulary system is a schedule that assists users in finding the appropriate *preferred term* from a term that is not part of the indexing language.[71] Terms that are not used for indexing (e.g., do not occur in object's descriptive surrogates) are described as *non-preferred terms* and may be part of an *entry vocabulary* that links frequently used words that users associate with their information needs to the concepts or categories used to classify objects and assign index terms (*preferred terms*) to objects.[72]

---

[71] Dagobert Soergel, *Indexing Languages and Thesauri* (Los Angeles, CA: Melville Publishing Company, 1974); Dagobert Soergel, *Organizing Information: Principles of Data Base and Retrieval Systems,* (Orlando, FL: Academic Press, Inc., 1985).

[72] The methodology for constructing an indexing language is treated extensively by Soergel, including methods for treating morphological variants, spelling variants, consolidating synonyms, and consolidating quasi-synonyms to construct a set of descriptors to be used in indexing. The intention of this part of

In a traditional controlled vocabulary system, a lexicographer selects a preferred term to act as the preferred form for a generalized concept formed by combining *synonymous terms* (ST) and terms that are functionally equivalent from a system perspective known as *equivalent terms* (ET). Where a preferred term is used to describe a concept that subsumes one or more preferred terms, the term is a *broader term* (BT). The term subsumed is a *narrower term* (NT).

The information structure represented by an indexing language that contains broader terms, narrower terms and equivalent terms is a hierarchical structure. Broader terms can be conceptualized as a generalization formed by conceptually combining two or more narrower terms. For example, one might define *physics* to subsume the concepts individually represented by *quantum mechanics* and *motion*. A feature can be assigned to represent a broader term whose truth-value is a function of a disjunction in which one of a set of alternatives is *true*, for example (*physics* = *true* IFF(*quantum mechanics* = *true* or *motion* = *true*))[73]. In this type of situation, a feature's truth-value is inferred from the truth-values of two or more features within a single case that represent narrower terms applicable to that case.

---

the paper is to provide a conceptual basis for understanding how to conceptualize a representation in deterministic inductive logic. The details associated with the construction of indexing languages and thesauri are peripheral to and beyond the scope of this research.

[73] IFF is used to represent the condition *if and only if*.

Deterministic inductive logic facilitates defining a broader term by combining two or more case vectors, where the collection of cases is used to define a vector that is used to represent a concept (broader term) as a category of cases whose members are representatives of the concept captured by the broader term. The effect of combination is to create a category of cases that are individually representatives and collectively frame the scope of the concept identified by the broader term.

The feature vector formed by the deterministic inductive combination of a set of case vectors can be used as a conditional for adding a feature to the feature vector representation. The new feature will have a truth-value of *true* when the conditions match those expressed by the combined feature vector and a value of *false* for cases that do not meet the necessary criteria. The effect is to create a new feature, based upon the discriminants of a category that can be used in conjunction with existing or new cases in conjunction with the test for subsumption (see Figure 15) to record the new concept.

The two approaches for defining broader terms are very different. The traditional approach for creating pre-combined descriptors is accomplished using a deductive inference while the second accomplished with DIL combination applies induction from examples. The first approach is typical of existing controlled vocabulary systems that are based upon deductive logic models. The second approach is only possible with induction.

It is possible that a deterministic inductive combination of cases could result in a feature vector that contains only *indeterminate* values. In such situations, a feature may be added and assigned to all the selected example cases as was described in Chapter 3 in conjunction with the description of defining features by enumeration. This approach results in *true* being assigned to that feature for selected applicable cases and *false* assigned to all other cases. This type of feature may be useful pragmatically to a specific user or user group; however, since the feature's assignment is dependent upon human intervention and identification of applicable cases instead of systemic assignment it cannot be assigned automatically.

Equivalence relationships (ET) can be assessed by comparing two vectors and determining the degrees of similarity (DOS(COMPARE(a, b)) = 1.00) or difference (DOD(COMPARE(a, b)) = 0.00).

Hierarchical relationships (BT and NT) can be assessed using the rules for testing for hierarchal subsumption (See Figure 15).

In some controlled vocabulary-based systems, *related terms* (RT) are associated where two or more terms share some conceptual association. Generally the goal of a related term is to make a user aware of a related concept that they might consider in constructing a query. Related terms have been traditionally handled by references in thesauri that link multiple terms. One way of conceptualizing two related terms is by defining a broader term that represents their association. For example, a *development date* may be related to a *deployment date*, where the *development date* should logically precede the *deployment date*. If these two dates are related in conceptualizing a system's lifecycle, an alternative to defining *related terms* is to collocate them and associate both with a *broader term*. Since objects are represented in deterministic inductive logic with feature vectors, one should define *related terms* by associating them with a *broader term* and by adding a new feature to the feature vector for recording the broader term. The truth-value *true* would then be assigned to every case for which any of the related terms has a value *true* and *false* when no related term has a value of *true*.

Recall, precision, discrimination and fallout

The pragmatic quality of any retrieval is a function of the appropriateness of the collection to the user's information needs, the quality of the formulation of those needs, and the ability of the system to represent the concepts that are relevant to discriminating documents that are appropriate to the context formed by the characteristics of the user and their information needs.

Relevant documents are those that are appropriate, while non-relevant documents are not appropriate. Appropriateness can be based upon many factors, including topicality, originality, uniqueness, language accessibility to the user, reading level, etc., as each possible factor relates to the context of the user and their information needs. The determination of whether an object is relevant, or its degree of relevance, or whether its relevance is based upon the object's pertinence or utility to the user—information need context has been the subject of extensive research. The nuances of relevance research are not essential to deterministic inductive logic, however, they should be considered if deterministic inductive logic is applied to building a system to fulfill user's information needs.

Four metrics are commonly used to evaluate retrieval and filtering systems' performance: *recall, precision, discrimination* and *fallout*. The overall quality of a retrieval set is commonly measured by the metrics *recall* and *precision*. *Recall* is an estimate of the proportion of relevant documents retrieved of all the relevant documents in a collection. *Precision* is a measure of the number of relevant documents as a percentage of the total number of documents in a retrieval set. *Discrimination* is a performance metric that addresses the ability of a system to correctly reject non-relevant documents. *Fallout*, is the complement of discrimination, a performance metric that reflects the proportion of non-relevant documents retrieved as a percentage of all non-relevant documents in a collection.

All four measures are reliant upon relevance judgments and relevance information, which is typically not available in actual operational situations. Assessing relevance is a problem that confounds most information filtering and retrieval research. Specific values of each metric are dependent upon the availability and quality of relevance information. Relevance information is dependent upon many variables and thus the interpretation of performance metrics into a meaningful system performance appraisal, requires addressing the issues of relevance more extensively than is necessary to show how these metrics relate to deterministic deductive and inductive system designs.

Of the four measures; recall, precision and fallout are proportionalities that are based upon the characteristics of both retrieval sets and the collections from which they were drawn. Discrimination, on the other hand, can be applied on a case-by-case basis by examining the characteristics of the representations of the query and the document. If the characteristics that make a document *not relevant* are not part of a system's vocabulary, it is clearly not possible to formulate a query that specifies precisely which documents to exclude, because the characteristics of the exclusionary rule required to correctly reject documents cannot be represented by that system. From a complementary perspective, if the characteristics that make a document *relevant* are not part of the system's vocabulary, it is not possible to formulate a query that would precisely include relevant documents because the characteristics of the inclusionary rule required to correctly select relevant documents cannot be represented in the system.

Discrimination, in a vocabulary-based system, is a function of the ability of a system's vocabulary to represent or express the concepts that would either cause an object to be correctly rejected or correctly accepted by a retrieval or filtering system. Theoretically a system using a perfect query formulation should be capable of retrieving all relevant documents (100% recall) and only relevant documents (100% precision) if that system's vocabulary is sufficiently expressive to be capable of encoding the information essential to correctly discriminate the relevance of every document in a collection with respect to the perfectly formulated query. To achieve perfect discrimination, a system's vocabulary, must be sufficient to represent every concept or characteristic that would cause a document to be *not relevant* (exclusionary rules) and every concept or characteristic that would cause a document to be *relevant* (inclusionary rules). If any of the inclusionary concepts overlap the exclusionary concepts used to describe the information need, the overlap will result in a non-deterministic relevance outcome, because the specificity of terms is not sufficiently precise. Therefore, for the system to achieve "perfect performance" there cannot be any overlap between the concepts used to include and the concepts used to exclude documents.

The process of increasing the scope of retrieval should result in an increase in recall and ordinarily results in a concomitant decrease in precision. Broader terms are used to facilitate recall strategies in vocabulary-based systems. The creation of broader terms has previously been treated. Table 29 illustrates defining a broader term using an enumerated list of cases.

Table 29. Defining Broader Terms by Combination

	Gin	Vodka	Bourbon	Scotch	Spirits
[A] Gin	T	F	F	F	T
[B] Vodka	F	T	F	F	T
[C] Bourbon	F	F	T	F	T
[D] Scotch	F	F	F	T	T
A or B or C or D	I	I	I	I	T

In the example illustrated in Table 29, if users were interested in retrieving (Gin or Vodka) they could search for NOT(Scotch or Bourbon) and achieve the same objective. This is an example of equivalent queries.

The process of decreasing the scope of retrieval in vocabulary-based systems ordinarily results in an increase in precision and a decrease in recall. This occurs in part because, for economic reasons, objects are assigned a small number of specific descriptors. It is not reasonable to expect, in any system with a limitation on the number of descriptors that can be assigned, a descriptor for every possible retrieval objective. The experimentally observed reduction in recall associated with increased precision would not theoretically occur in a system with unlimited descriptors and optimal query formulation.

The example in Table 30 demonstrates the traditional trade-offs in improving precision in systems without unlimited descriptors.

One can improve recall of specific toys by using general categories of toys such as “soft toys” and “toys” in the example in Table 30. If, however, one were to refine a search of “soft toys” without a feature to accommodate the “stuffed camel” one would miss that object with a search that specified either the “teddy bear” or the “foam ball.” One benefit of an inductive model is that any user can extend a representation at any time to accommodate new or desirable concepts.

Table 30. Indexing Toys Example 1

	Teddy Bear	Marble	Foam ball	Soft toys	Toys
Stuffed camel	F	F	F	T	T
Foam ball	F	F	T	T	T
Baseball bat	F	F	F	F	T
Marble	F	T	F	F	T

Table 31. Indexing Toys Example 2

	Teddy Bear	Marble	Foam ball	Soft toys	Toys	Stuffed camel	Basebal l bat	Rigid toys
Stuffed camel	F	F	F	T	T	T	F	F
Foam ball	F	F	T	T	T	F	F	F
Baseball bat	F	F	F	F	T	F	T	T
Marble	F	T	F	F	T	F	F	T

Adding a feature to the end of the feature vector extends the representation. In the example in Table 31, a feature was added to represent the two distinct objects “stuffed camel” and “baseball bat,” and a general category was added to differentiate between “soft” and “rigid” toys. In the last line of Table 31, the effect of the representation’s changes to the “marble” object demonstrates that by adding the category rigid toys, we have increased the number of ways of retrieving marbles.

A change in a system's representation requires retrospective indexing. In traditional vocabulary-based systems retrospective indexing is generally not attempted for economic reasons. In the deterministic inductive model one can retrospectively index in two ways, (1) by employing the *unknown* truth-value, or (2) by implication from existing features.

Table 32. Extending Feature Vectors with Unknown Truth-Values

	Teddy Bear	Marble	Foam ball	Soft toys	Toys	Stuffed camel	Baseball bat	Rigid toys
Teddy Bear	T	F	F	T	T	U	U	U

In Table 32, the representation of the teddy bear object is extended using the *unknown* truth-value. This does not detrimentally affect the retrieval of that object using the representation illustrated in Table 30, before it was extended in Table 31, however, it obviously does not allow a user to include or exclude the teddy bear object using any of the three new features.

In the example in Table 33, the representation of a teddy bear object is extended by using two conditionals (1) IF ("soft toys" = *true* THEN ("rigid toys" = *false*) and (2) IF ("rigid toys" = *false*) THEN ("baseball bat" = *false*).

Table 33. Representing a Teddy Bear with Extended Features

	Teddy Bear	Marble	Foam ball	Soft toys	Toys	Stuffed camel	Baseball bat	Rigid toys
Teddy Bear	T	F	F	T	T	U	F	F

There is no benefit to encoding the “stuffed camel” feature for any object that is not a stuffed camel, unless one anticipates requiring that feature to formulate an exclusionary rule. If for example, a user were to retrieve a set based upon the inclusionary criteria “soft toys” and wished to exclude “stuffed camel” both the “teddy bear” and “foam ball” would have to be examined to assess whether they were or were not “stuffed camels.” If such a situation were desirable, in an inductive system, updating the truth-values for the extended features might capture a user’s analytical efforts. This evolutionary characteristic of an inductive system might make retrospective indexing possible in some situations by opportunistically capturing information during user interactions with a system. A user’s analysis could be captured for later re-use, or so that other users might benefit from the knowledge discovered by a previous user interacting with a system.

Encoding structured information in feature vectors

Structured information is ordinarily referenced by a link between an identifier of an information element and a storage space in which a value, within some range of acceptable or expected values, is to be stored and from which a value will be retrieved. For example, a two-digit field that is used to record the month portion of a date, which is limited to the range $1 \leq x \leq 12$, could be identified with a name that can be used to reference a two-byte storage location. If one were constructing a system in which twelve months of information were reconciled during a closing period, the “month” range might be extended to thirteen ($1 \leq x \leq 13$) to accommodate references to a closing period. If the identifier were *employment-starting-date-month*, the symbol might refer to a two-byte storage area in which an appropriate integer could be stored.

Generally, field values can be described in terms of either a quantitative comparison {<, ≤, =, ≥, >} or a logical comparison {equivalent, true, false} where the comparison is made between two values, formulas or statements. Quantitative comparisons may require formulas that take advantage of arithmetic operators {+, -, *, /, ^} and logical comparisons may require formulations that take advantage of logical operators {AND, OR, NOT}. Computer languages, such as Perl,[74] that are designed to manage structured information ordinarily support all of these capabilities.[75]

---

[74] Larry Wall, Tom Christiansen, and Randal L. Schwartz, *Programming Perl*, 2d ed. (Sebastopol, CA: O’Reilly & Associates, Inc., 1996).

[75] There may additionally be string comparison functions that return a logical truth-value for various string comparisons. Perl provides excellent support for data manipulation and is an example of a language with excellent support for converting information values into logical truth-values. Details of all the possible

Any statement that can be evaluated to be either *true* or *false* from a Boolean perspective can be used to define a feature. Table 34 contains examples of valid approaches for defining features in deterministic inductive logic.

Table 34. Examples of Mapping Quantitative Conditionals to Logical Features

Conditional formulation	Value resulting from conditional evaluation
X=3: 0 < x < 12	*true*
a=2, b=3: a + b < 25	*true*
a=2, b=3: a^2 + b^2 < (a + b)^2	*true*
a=1, b=1: a^2 + b^2 < (a + b)	*false*

A structured information element can be mapped to one or more features, for example, a value for income can be converted to a faceted classification by defining ranges of income and assigning either a binary truth-value or one of a list of values formed as an *n-facet* feature. Table 35 illustrates how features can be assigned facet based truth-values or Boolean truth-values given ranges of quantitative values:

Structured text information can be mapped to features using comparisons in the same way that numeric information is mapped. For example, a string comparison can be used to determine whether a field contains the text "Brewer" where the result of the comparison is used to assign a truth-value for the presence

---

approaches are beyond the scope required to understand the relevant concepts in this chapter.

or absence of that string within a field. Similarly a system could convert entries in a field used to represent "toy type" and assign feature values based upon the field's contents. If, for example, the *toy-type-field* contained the value "marble" the feature that records whether an object is a marble in our toy example could be set to the value *true*. Conversely, if the value of *toy-type-field* was not "marble," the truth-value *false* could be assigned to the feature used to represent whether the toy is a marble.

Table 35. Mapping Quantitative Values to Faceted Features

Value	$0 \leq x \leq$ 1,000 [A]	$1000 < x \leq$ 5,000 [B]	$5000 < x \leq$ 10,000 [C]	$10{,}000 < x \leq$ 25,000 [D]	$x >$ 25,000 [E]	Faceted feature
14,356	F	F	F	T	F	D
537	T	F	F	F	F	A
3976	F	T	F	F	F	B

An assessment of the different ways one might represent a specific feature, as defined by the values expressed in a field, may be required to construct a representation that correctly maps structured information elements to features in a way that supports all desired comparisons between and among features expressed by user-meaningful groupings of cases. If, for example, the values "marbles" "marble" and "MARble" are all to be interpreted as indicative of asserting a *true* truth-value for the feature *marble*, it may be necessary to account for spelling and other morphological variants to correctly map the various expressions to correct feature values.

If the individual values of objects in structured information elements are classified into categories, they can be mapped to faceted features. It is possible to map many valued fields to features where each possible value is mapped to a respective feature, however, the practicality of that is a function of both the magnitude of different values and the facility it contributes to formulating and reasoning about user-meaningful aggregates.

Where the goal is to inductively discover the classes that are expressed in a database, as one might do in data mining, there may be benefits to mapping every possible value to a different feature and then to rely on the deterministic inductive combine function to discover how those unique values are expressed in user-meaningful groups.

Chapter V – Representational sufficiency in deterministic inductive logic

A system's vocabulary is *sufficient* with respect to a specific *user-information need* if it is possible to express the relevance conditions that a user requires to formulate a perfect query that, if executed would result in perfect performance (100% recall, 100% precision). It is not important to the definition whether any particular user can formulate any particular perfect query or whether any user would be willing to expend the effort required to refine a query to the level of a perfect formulation. It is likewise not important to the definition of sufficiency that we account for the effects of information presentation and use on changes in a user's perspectives of relevance. The concept of representational sufficiency is theoretical and is present whenever it is possible to formulate a perfect query in a system vocabulary that achieves perfect performance.

A perfect query contains two parts; *inclusionary discriminants* and *exclusionary discriminants*. *Inclusionary discriminants* are concepts represented in a system's vocabulary that are always present in relevant documents. *Exclusionary discriminants* are concepts represented in a system's vocabulary that are never present in relevant documents. Concepts that are ambiguous with respect to a specific *user-information need* are terms in a system's vocabulary that may be present in relevant and non-relevant documents. They are not *discriminants*, because they hold a truth-value that is *indeterminate* with respect to assessing the relevance of a document given a specific *user-information need* context.

When there are one or more *discriminants* that can be formulated into a set of inclusionary and/or exclusionary rules that collectively completely and accurately describe the scope of relevant objects for a given information need, the vocabulary is *sufficient* to express that need. When a vocabulary doesn't contain the required terms (insufficient exhaustivity) or the terms are not specific (insufficient specificity) then the vocabulary is not sufficient to meet the specific need. *Vocabulary sufficiency* addresses whether a system's vocabulary can accurately, completely and correctly represent a user's need such that a query formulation can be executed which will result in all and only relevant documents.

One can identify discriminants with deterministic inductive logic by combining two or more relevant objects and comparing the aggregate representation with the identity-vector. All of the features that are always true will have a truth-value of *true.* Additionally, one can contrast relevant and non-relevant objects to assess which characteristics are not shared by relevant and non-relevant objects. The combine, compare and contrast functions of deterministic inductive logic facilitate identifying the discriminants of a *user-information need* by providing a user with system tools for finding discriminants from examples of relevant and non-relevant objects.

Not all information storage and retrieval systems represent objects with vocabularies that are based upon languages used for human communications. A quantitative term frequency distribution, based upon the proportion of the frequencies of words in a text to their frequency in a corpus, is a commonly used approach to representing documents in many modern information filtering and retrieval systems. While this research does not address any of the quantitative representation models, but focuses instead on the vocabulary-based model, any characteristics that can be used to represent objects can be used to determine the truth-value of a feature in deterministic inductive logic. Since it is possible for a feature's truth-value to be based on the result of a statistical classifier or some other quantitative model, for the purposes of this chapter the terms representation and vocabulary are used interchangeably to represent equivalent concepts.

Representational sufficiency is not limited to the availability of terms in an indexing language. The concept, while possibly easier to describe in terms of a system's *vocabulary* is applicable irrespective of how a feature is defined, provided there is a method that consistently assigns truth-values to that feature.

In the Logan-Bonnie group example (see Table 11), black Odyssey and Passport automobiles were the automobiles appropriate to the market segment defined by the preferences of Logan and Bonnie. By combining Logan-Bonnie and contrasting with a difference vector, the exclusionary rule (~Accord, ~Civic, ~CRV, ~Blue, ~Red, ~Silver, ~Green) was determined. In that case Black, Odyssey and Passport were all ambiguous terms used to represent features that could occur in automobiles that were appropriate to the defined market segment and to automobiles that would be appropriately excluded from the market segment defined by Logan and Bonnie.

In the toy example, it was necessary to extend the initial representation to retrieve all and only “stuffed camel” or “baseball bat” objects. While it is possible to retrieve both types of objects using the “toy” feature in an inclusionary rule, the resulting set will also contain objects that are not appropriate.

Table 36. Indexing Toys with Limited Features

	Teddy Bear	Marble	Foam ball	Soft toys	Toys
Stuffed camel	F	F	F	T	T
Foam ball	F	F	T	T	T
Baseball bat	F	F	F	F	T
Marble	F	T	F	F	T

Adding three features, assigning truth-values using inference rules, and filling all remaining features with the *unknown* truth-value results in a representation that is sufficient to discriminate some objects but insufficient to discriminate others. The extension results in a sufficient feature set but without all the necessary truth-values that must be known for the system to be sufficient to discriminate every possible retrieval objective.

As a practical matter, the *unknown* values can be used to guide the analysis of an information system and to add knowledge to a knowledge base represented using the deterministic inductive logic paradigm.

Table 37. Indexing Toys after Extending the Feature Set

	Teddy Bear	Marble	Foam ball	Soft toys	Toys	Stuffed camel	Baseball bat	Rigid toys
Teddy Bear	T	F	F	T	T	U	F	F
Foam ball	F	F	T	T	T	U	F	F

A material advantage of inductive systems is that they can "learn" and be updated during their lifespan of use. Systems designed using the deductive model are fixed during the user's information search process. The deterministic inductive model allows extending a representation without degradation to the existing system and provides the necessary facilities for adapting a knowledgebase to reflect the learning that has been demonstrated to occur during a user's information search process.

Representational sufficiency is a function of the *user-information need* and user's interests in discriminating different objects. Representational sufficiency is not inherently a function of the characteristics of objects. The ability to discriminate every possible combination of objects is not necessary to accomplish a specific retrieval objective. To satisfactorily accomplish a retrieval objective only requires whatever representation is necessary to discriminate object's that are relevant from ones that are not relevant within the constraints of a specific information need.

The requirements for representational sufficiency are a function of the ways objects are compared and contrasted. Table 38 contains ten situations that contain different configurations of objects that will be used for illustration in the remainder of this chapter. Each configuration of objects requires a different representation depending upon what differences are of importance to the user.

Table 38. Object configurations

Object description
[1] marble in a cardboard box
[2] marble in a brown cardboard box
[3] brown marble in a brown tin box
[4] brown marble in a cardboard box
[5] brown marble in a green cardboard box
[6] green marble in a green cardboard box
[7] green and brown marbles in a cardboard box
[8] green and brown marble in a cardboard box
[9] marble in a green and brown cardboard box
[10] marbles in green and brown boxes

In the Table 39, combinations of the ten situations are contrasted. The groups that must be differentiated are listed in the first column. The characteristics required to discriminate those differences are described in the second column. In the third column a set of features that could be used to record those characteristics is identified. The illustration is additive, and so each row should be interpreted as an additional relevant differentiation. The feature set in column three that is required to differentiate the condition in a specific row is the set of features listed in column three of that row and some or all of the features identified in column three's predecessor rows.

Table 39. Differentiating Different Configurations of the Objects in Table 38

Differentiate	Characteristic of difference	Needed feature extensions
[10] from [1,2,3,4,5,6,7,8,9]	Boxes versus box	F(1)-Singular box F(2)-Multiple boxes
[1] from [2]	Cardboard box versus brown cardboard box	F(3)-brown box
[2] from [3]	(a) Cardboard box versus tin box (b) brown marble versus unknown-color marble	F(4)-cardboard box F(5)-tin box F(6)-brown marble
[1] from [2] from [3]	The color of the marble (brown) is the characteristic that differentiates [1] from [3]. The other differences are the same as those defined in [1] versus [2] and [2] versus [3]	F(6)-brown marble
[4] from [8]	Brown marble versus green and brown marble	F(7)-green and brown marble

Table 39 illustrates the following characteristics of representational sufficiency:

- If the only case that is of interest is the one in which there is more than one box, there is no need to represent marbles or colors to achieve representational sufficiency.

- When differentiating between “marble in a cardboard box” and “marble in a brown cardboard box” both share the characteristics “marble in a box” and so the difference required to achieve representational sufficiency is a feature to record the color of the box.

If the differences in the preceding example are conceptualized as cumulative, the seven features selected by contrasting the five situations provides a representation that is sufficient to discriminate the types of cases illustrated by Table 39.

If, instead of building the representation incrementally, one were to capture all of the situation characteristics, the sufficient feature set could have been derived using a contrast function to assess relevant differences. Tables 40 and 41 illustrate the relevant features necessary to differentiate the first, second and third object configurations from Table 38.

Table 40. Identifying Relevant Discriminants (Features 1-6)

	marble Sing.	marbles Plural	box Sing.	boxes Plural	Tin box	Card board box
[1]	T	F	T	F	F	T
[2]	T	F	T	F	F	T
[3]	T	F	T	F	T	F
Contrast([1],[2])	F	F	F	F	F	F
Contrast([2] [3])	F	F	F	F	T	T

If the representation is designed to take advantage of the facet capability of deterministic inductive logic, it could be represented using the seven facets in the table 42.

Table 41. Identifying Relevant Discriminants (Features 7-10)

	Green marble	Brown marble	Brown box	Green box
[1]	U	U	U	U
[2]	U	U	T	F
[3]	F	T	T	F
Contrast([1],[2])	F	F	T	T
Contrast([2] [3])	T	T	F	F

Table 42. Organizing the Features into a Sufficient Set of Facets

Feature (1) marble, marbles [facet]	A = marble (singular) B = marbles (plural)
Feature (2) box, boxes [facet]	A = box (singular) B =boxes (plural)
Feature (3) box type [facet]	A = tin B = cardboard *unknown*
Feature (4) green color marble(s)	*true* *false* *unknown*
Feature (5) brown color marble(s)	*true* *false* *unknown*
Feature (6) green color box(es)	*true* *false* *unknown*
Feature (7) brown color box(es)	*true* *false* *unknown*

The representation of the three distinct situations in Table 43 and the needed facets to discriminate these situations is illustrated in Table 44.

Table 43. Selected Object Configurations

[1] marble in a cardboard box
[2] marble in a brown cardboard box
[3] brown marble in a brown tin box

Table 44. Representation Required to Discriminate Objects

	F(1)	F(2)	F(3)	F(4)	F(5)	F(6)	F(7)
[1]	A	A	B	U	U	U	U
[2]	A	A	B	U	U	F	T
[3]	A	A	A	F	T	F	T
Contrast([1], [2])	F	F	F	F	F	T	T
Contrast([2], [3])	F	F	T	T	T	F	F
Contrast([1], [2],[3])	F	F	T	T	T	T	T

In the faceted model, features one and two are not required to differentiate the first three situations, while features four, five, six and seven are required to differentiate those distinct situations.

The features required to achieve representational sufficiency are directly related to how the feature vector is designed and how features are used to represent object characteristics. Using faceted features may reduce the total number of features required in a representation.

Chapter VI. Application illustration: Document assembly problem

This chapter will illustrate the application of deterministic inductive logic to compiling documents using a case study. Specifically this chapter looks at the problems of solicitation and contract document formation under the Federal Acquisition Regulations (FAR)[76] and illustrates the advantages of using deterministic inductive logic in designing a system to represent and select standard texts for inclusion in instruments in an environment in which changes may continuously occur both in the laws and regulations controlling acquisitions in general and in the needs and relationships between suppliers and users in a specific acquisition action .

Overview

The FAR provides a regulatory framework for government agencies and organizations to manage their acquisitions. Acquisition management can be broken into a number of components from planning through final disposition and contract closeout. An acquisition process essentially begins with a user requirement that is formulated into a formal specification and combined with

[76] The Federal Acquisition Regulations are chapter 48 of the Code of Federal Regulations. The citation of a regulation such as part 52 within the FAR is properly cited as 48 CFR 52.000 which refers to Title 48, Part 52, section 000. Throughout the balance of this chapter, we refer to FAR sections using the shortened notation 52.000.

terms and conditions into a solicitation document that is used to guide the source identification, qualification and selection process. This is the solicitation phase of an acquisition. Source selection allows the acquisition to move into the contract phase during which products and services are designed, manufactured, assembled, packaged, delivered, tested, inspected, accepted and compensated.

The two controlling instruments that communicate the intentions and the terms and conditions for the interactions between the requesting agency and the market or specific supplier are the solicitation and contract.[77] The FAR provides guidelines for every aspect of an acquisition action and for the management of the acquisition process. The FAR is organized into 53 parts. The first 51 parts describe in detail the regulatory advice and guidance for handling federal acquisitions and circumstances in which acquisitions are conducted. Part 52 contains standard text elements that are used in acquisition instruments and part 53 provides standard forms.

A solicitation is used to request information from bidders and offerors. This information is used in the selection of a source to fulfill an acquisition requirement. A contract is an instrument that commits an agency and a source to a process in which an acquisition requirement will be fulfilled. The standard text

---

[77] Interagency agreements are handled using Memorandums of Agreement (MOA), which may or may not also be associated with solicitations and contracts. We have elected not to address the formation of an MOA document, as they are both infrequent and very context specific.

elements used in these instruments are known respectively as *solicitation provisions* and *contract clauses*. The basic standard text may, under certain circumstances, include a requirement for information to be filled in that is specific to the contract action and to the provision or clause that is to be modified to accommodate a specific objective.[78]

In constructing an acquisition instrument, one selects the provisions and clauses that, given the specific characteristics of that acquisition, are either legally required in all instruments (*mandatory*), are required in all instruments of a specific type (*required when applicable*), or that are deemed by the Contracting Officer (CO) to be appropriate for conducting the specific acquisition (*optional*). These three conditions can be conceptualized in terms of two distinct classes of decisions. The first class includes all decisions that result from comparing the characteristics of the acquisition action with the characteristics under which the action is conducted without respect to considerations of the CO's preferences. The second class addresses preferences that the CO may make to achieve the best overall value for the government.

Management and control over acquisitions involves multiple branches of government and multiple levels of executive authority. The financial resources

---

[78] An "alternate" is a substantive variation of a basic provision or clause prescribed for use in a defined circumstance. It (1) adds wording to, (2) deletes wording from, or (3) substitutes specified wording for a portion of the basic provision or clause as changed by the addition, deletion, or substitution (see 52.105(a)). 48 CFR 52.101 Using Part 52.

used to pay for acquisitions are the result of appropriations, which are within the purview of the Legislative branch of the U.S. government. The expenditure of funds to accomplish specific acquisition objectives is done by and under the control of the Executive branch of the U.S. government. The authority of the head of contracting for an agency is derived from the authority of the head of the agency. The personnel engaged in acquisition activities all derive their authority by delegation from the head of contracting of an agency through however many levels of organization are required to reach the level at which the acquisition activity is being conducted.[79] Figure 19 illustrates an example of the delegation of authority. Each organizational level may restrict the authority of a lower organizational level and thereby reduce the alternatives a CO might otherwise be allowed by regulation.

The Department of the Treasury may elect to reduce the flexibility available to a CO from that described in the FAR to reflect agency selections. For example, the Department of Treasury may elect to make incorporating a Value Engineering program[80] a requirement whenever it might otherwise be at the

---

[79] In certain specialized situations the authority to enter into an acquisition may require a delegation from outside of the agency seeking to acquire products or services. For example, in the case of acquiring federal information processing resources, a delegation of authority is required from the General Services Administration.

[80] Value Engineering is a program intended to seek contractor assistance in the identification of cost effective improvements in products and services acquired by the government. See FAR Section 48.

discretion of the contracting officer for acquisitions conducted by the IRS but not ones conducted by FLETC. The *acquisition context* is defined by statute and regulations to control the range of acceptable choices and actions permitted to acquisition management personnel engaged in specific acquisition actions. The characteristics of a specific acquisition action and the choices elected by the CO responsible for a specific action define the *acquisition action context.*

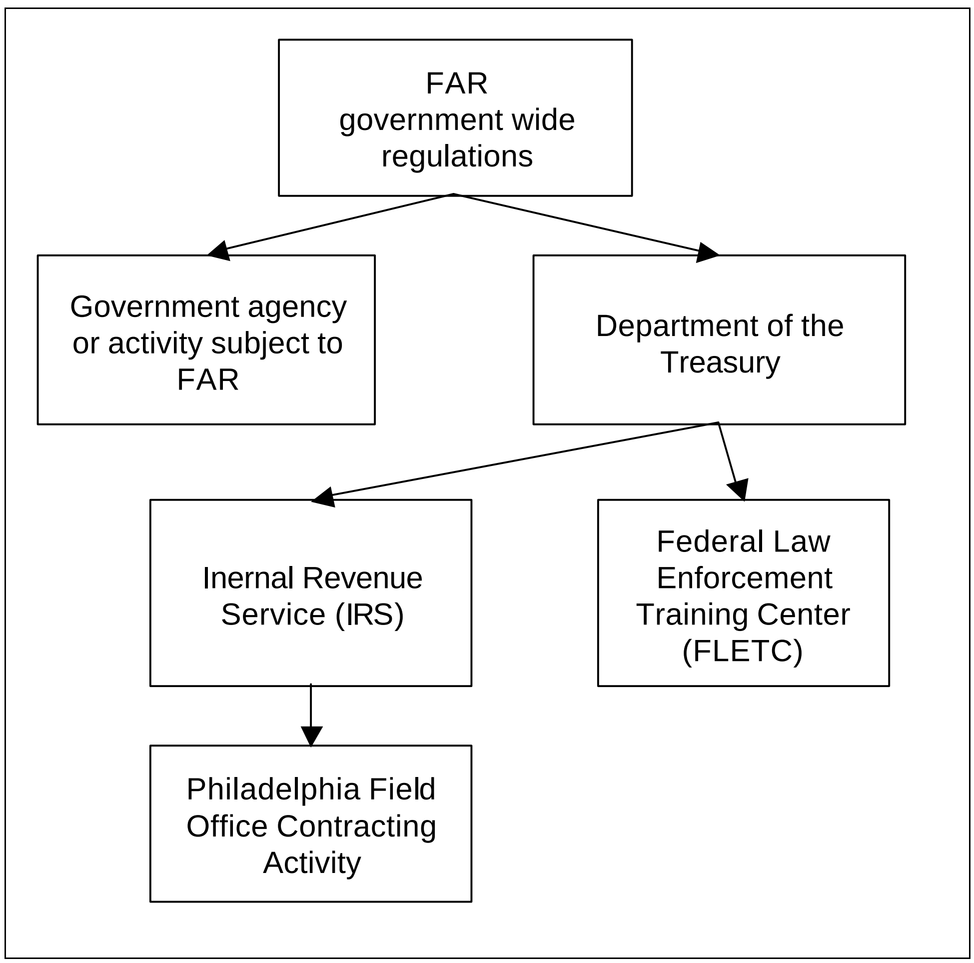

Figure 19. Levels of contracting authority

We can design a general approach to constructing documents that is applicable to both solicitation and contract instruments, by constructing a database of the texts that are includable in any acquisition instrument and providing a mechanism for selecting the appropriate or relevant texts in accordance with the rules for the applicability and use of provisions and clauses. This design is based on encoding the decision criteria that control provision and clause selection and providing a capability for matching those rules with the requirements of a specific acquisition. This process can be conceptualized as a retrieval or filtering operation in which the texts incorporated in part 52 of the FAR, given their prescribed applicability and use as documented in FAR parts 1 through 51, are matched to a profile that is defined by the characteristics of a specific acquisition.

As a practical matter the characteristics of an acquisition may evolve throughout the acquisition process. Particularly in complex acquisitions, it is quite common for decisions made early in the acquisition's lifecycle to be modified by new information. Additionally, there are some decisions that a CO must make to qualify a contract clause that cannot be made until discussions are conducted with responsible offerors. For these reasons, the design must accommodate a selection process that is both non-monotonic (accommodates decision revision) and supports unknown values for decisions that cannot be made with the information available at a particular point during the acquisition process.

Deterministic inductive logic directly accommodates unknowns with the truth-value *unknown.* An iterative decision revision process, in which a user is permitted to modify the acquisition action context and then cycle through the applicability and use rules to select or re-select text components from the standard text database of provisions and clauses, facilitates the dynamic adaptation of acquisition instruments to accommodate the unique characteristics and conditions presented by specific acquisition actions.

The compare function of deterministic inductive logic can be used to test for any changes in the acquisition action context by recording the conditions prior to modification in a vector (*before_vector*) separate from the vector used to capture incremental changes and decision revisions (*after_vector*). If the feature vectors are identical, the selection of provisions and clauses will not require re-filtering, however, if a change has occurred (DOS(compare(*before_vector*, *after_vector*)) < 1.00), provisions or clauses may need to be either added or deleted depending upon the user's specific feature value selections and the rules for applicability and use of provisions and clauses.

The provisions and clauses that are mandatory, required when applicable, or optional for a specific acquisition action are determined by the interaction of the *acquisition context* and the *acquisition action context.* The application of deterministic inductive logic to the construction of acquisition instruments under FAR can be separated into the two classes (1) mandatory or required when applicable, and (2) optional.

Mandatory or required when applicable provisions and clauses arise from decisions that result from comparing the characteristics of the acquisition action with the characteristics under which the action is conducted where, for whatever reason, the contracting officer's preference is not required to resolve whether a provision or clause is or is not applicable to an acquisition action.

Optional provisions and clauses arise from decisions in which the FAR permits the head of a contracting office or a CO to elect to include or exclude a provision for a specific acquisition action where that preference is intended to facilitate achieving the best overall value for the government.

Defining feature vectors to represent provisions and clauses in the acquisition context

The construction of a feature vector sufficient to facilitate the selection of provisions and clauses requires examining every rule for applicability and use and identifying the features that discriminate each provision's and clause's selection. For example, the differentiation between whether a standard text object is used in a solicitation or a contract or both can be handled by creating a set of features to represent the appropriate use of each standard text object.[81]

---

[81] It is common accepted practice to incorporate the solicitation and whatever response a bidder submitted in response to that solicitation (quote or proposal) by reference in a contract. From that perspective, one might argue that all solicitation provisions are incorporated by reference in contracts. Similarly, a component of a solicitation is the anticipate method of procurement and the terms and conditions that are expected to be included in any contract that results from a

From the perspective of deterministic inductive logic there is no conceptual difference between the *mandatory* and a *required when applicable* types of conditions, in that each will be encoded with a truth-value equivalent to *true* in a feature that is applicable in a specific acquisition. For example, a provision that is mandatory in all solicitations would be encoded with a *true* in feature(1) of Table 45 *type-of-instrument-solicitation*. Similarly, a clause that is mandatory in all contracts would be encoded with a *true* in feature(2) of Table 45 *type-of-instrument-contract*. A provision is not included in a contract and so all solicitation provisions should have feature(2) *type-of-instrument-contract* encoded

Table 45. Features to Represent Types of Instruments

Feature identifier	Feature name	Applicability and use
Feature (1)	*type-of-instrument-solicitation*	*true*—text object is used in solicitations *false*—text object is not used in solicitations
Feature (2)	*type-of-instrument-contract*	*true*—text object is used in contracts *false*—text object is not used in contracts

---

solicitation. From that perspective, all contract clauses could occur in a solicitation instrument as proforma contract terms and conditions. The representation example addresses encoding how text objects are prescribed for use in solicitations and contracts. Since text objects might be observed in either instrument retrospectively we have elected not to treat the instrument types (solicitations and contracts) as facets since they are not mutually exclusive throughout the acquisition process.

with a *false*.   Similarly, clauses are not included in solicitations and should therefore have feature(1) *type-of-instrument-solicitation* encoded with a *false*.[82]

The process of defining and implementing deterministic inductive logic feature vectors will be illustrated first for a single clause and then for a small set of provisions.

Accommodating the Definitions clause

Figure 20 illustrates how the FAR prescribes the applicability and use of a standard text. The specific text, 52.202-1 Definitions, in the example is used in acquisition instruments to define terms.

This specific example includes a number of features that would be required to control the inclusion or exclusion of the definitions clause. The required features are identified in Table 46.

---

[82] For the purposes of instrument preparation, while a proposed contract is ordinarily included as a part of a solicitation, and a solicitation is incorporated by reference in a contract, the documents are conceptually separate and distinct documents used to communicate the terms and conditions of connecting but separate phases of the acquisition process. As a practical matter the two instruments are always treated separately.

2.201 Contract clause.

The contracting officer shall insert the clause at 52.2021, Definitions, in solicitations and contracts except when the contract is not expected to exceed the simplified acquisition threshold. If the contract is for personal services, construction, architect-engineer services, or dismantling, demolition, or removal of improvements, the contracting officer shall use the clause with its Alternate I. Additional definitions may be included, provided they are consistent with the clause and the FAR.

52.202-1 Definitions.
As prescribed in Subpart 2.2, insert the following clause:

Definitions (Oct 1995)
(a) "Head of the agency" (also called "agency head") or "Secretary" means the Secretary (or Attorney General, Administrator, Governor, Chairperson, or other chief official, as appropriate) of the agency, including any deputy or assistant chief official of the agency; and the term "authorized representative" means any person, persons, or board (other than the Contracting Officer) authorized to act for the head of the agency or Secretary.

[Section deleted from illustration]

(g) Except as otherwise provided in this contract, the term "subcontracts" includes, but is not limited to, purchase orders and changes and modifications to purchase orders under this contract.

(End of clause)

Alternate I (Apr 1984). If the contract is for personal services; construction; architect-engineer services; or dismantling, demolition, or removal of improvements, delete paragraph (c) of the basic clause.

Figure 20. Definitions clause

Table 46. Additional Features Required by the Definitions Clause

Feature	Feature name	Applicability and use
Feature (3)	*Simplified-acquisition*	*true*—value of the contract is not expected to exceed the threshold for simplified acquisitions[83] *false*— value of the contract is expected to exceed the threshold for simplified acquisitions
Feature (4)	*Personal-services*	*true*— acquisition action is for personal services *false*— acquisition action is not for personal services
Feature (5)	*construction*	*true*—acquisition action is for construction *false*— acquisition action is not for construction
Feature (6)	*architect-engineer services*	*true*— acquisition action is for architect-engineer services *false*— acquisition action is not for architect-engineer services
Feature (7)	*dismantling, demolition, or removal of improvements*	*true*— acquisition action is for dismantling, demolition, or removal of improvements *false*— acquisition action is not for dismantling, demolition, or removal of improvements

[83] The threshold for simplified acquisitions changes over time. During the time when this research was conducted, the threshold for simplified acquisitions was affected by the streamlining of acquisitions and by the implementation of electronic acquisition technologies that use the Internet and agency credit cards. The specific threshold could therefore be a function of the specific agency or agency subunit; however, as a general matter, the simplified acquisition process is intended to account for approximately 80% of user requirements that are small purchases and can benefit from reducing complexity and paperwork.

If an acquisition action was expected to have a value above the threshold for a simplified acquisition, and that acquisition was not for personal services; construction; architect-engineer services; or dismantling, demolition, or removal of improvements, the entire clause at 52-202-1 would be included in both the solicitation and contract instruments. The feature vectors that might be used to represent the conditions for inclusion and exclusion of the two forms of standard text (Definitions (October 1995) and Definitions (October 1995) Alternate 1) are illustrated in Table 47.

Table 47. Representation of Definitions Clause

	F(1)	F(2)	F(3)	F(4)	F(5)	F(6)	F(7)
52.202-1	T	T	F	F	F	F	F
52-202-1 Alternate 1	T	T	F	T	T	T	T
Contract Action	F	T	F	F	F	F	F

If the conditions of the acquisition were modified so that the solicitation conditions changed from [TFFFFFF] to [TFTFFFF] because the estimated size of the acquisition did fit within the simplified acquisition threshold, then

DOS(compare([TFFFFFF], [TFTFFFF])) < 1.00 and the instrument would need to be reformed without the definitions provision.[84]

If we combine the conditions in which some form of Definitions clause would be included in an acquisition instrument, the presence of a Definitions clause is deterministically included in all solicitations and contracts that are not simplified acquisitions. The context of the specific clause is dependent upon whether the acquisition is for personal services; construction; architect-engineer services; or dismantling, demolition, or removal of improvements. The presence of mandatory conditions is illustrated in Table 48 by the features F(1), F(2), F(3) of the combined conditions, in that those features prescribe the absolute inclusion or exclusion of some form of definitions clause. Features F(4), F(5), F(6), F(7) define contexts in which a specific form of the definitions clause is required when applicable.

The treatment of optional provisions and clauses are dependent upon an election by the CO that a provision or clause is applicable to a specific

[84] As a result of a request for quotations or request for proposals, and the subsequent receipt of bids, a CO might determine that the acquisition could be completed within the threshold for simplified acquisitions. In this event, as a practical matter, the treatment of simplified acquisitions would result in making an award under the rules for simplified acquisitions.

acquisition, provided that the clause is not otherwise precluded from use in the type of acquisition anticipated by a higher authority.[85] The election can be

Table 48. Mandatory Versus Required when Applicable Features

	F(1)	F(2)	F(3)	F(4)	F(5)	F(6)	F(7)
52.202-1	T	T	F	F	F	F	F
52-202-1 Alternate 1	T	T	F	T	T	T	T
Combine	T	T	F	I	I	I	I

handled by providing a feature to be set for each specific acquisition to record that an authority has elected or not elected to include a provision or clause in the acquisition instruments for that action.[86]

Accommodating a small set of solicitation provisions

As can be seen from the single clause example, the logic required to construct solicitation and contract instruments includes many factors. Some of

---

[85] An Agency Head may specify how certain optional conditions are to be interpreted for acquisitions in that agency where the FAR provides for one or more alternative treatments. Elections by Agency Heads become mandatory for Contracting Officers operating under a delegation of authority from that Agency Head.

[86] In an actual implementation, the feature vectors used to describe the use and applicability of provisions and clauses may be modified to reflect the authority level of a specific user. In this case a hierarchy is required in which certain authority levels may be permitted option elections that are precluded to lower tier acquisitions personnel. For example, the acquisitions management of the IRS might fix an election that was delegated by the Department of Treasury for all Contracting Officers conducting acquisitions on behalf of the IRS.

the factors are dependent upon other factors and some are independent. Deterministic inductive logic facilitates selection of standard text by facilitating the assessment of which provisions or clauses should be included based upon the context of a specific acquisition. The test determines whether, given the terms for applicability and use of provisions or clauses, those provisions and/or clauses are logically subsumed by the context defined by the specific acquisition action. The resulting instrument includes that subset of the entire provision and clause database that is applicable to the specific acquisition action context.

We will illustrate the application of deterministic inductive logic using the prescriptions for applicability and use at 14.201-6 for some of the solicitation provisions that can be included in Invitations for Bids (IFB). The provisions prescribed by this subsection "are limited to subjects that are general in nature, do not come under other subject areas of the FAR, and pertain to the preparation and submission of bids."[87]

Figure 21 illustrates the following four types of conditions:

- The six provisions (52.214-1, 52.214-3, 52.214-4, 52.214-5, 52.214-6, 52.214-7) are all dependent upon the instrument being an IFB and only dependent upon the type of instrument.
- The provisions 52.214-9 and 52.214-10 are co-dependent on both the instrument being an IFB and whether the acquisition is to acquire

construction. The provisions 52.214-9 and 52.214-10 are excluded from IFB's for construction.

- The provision at 52.214-13 is applicable only in situations in which the instrument is an IFB and the CO decides to authorize telegraphic bids.
- The provision 52.214-14 addresses the place of performance for the contract action that will result from the IFB.

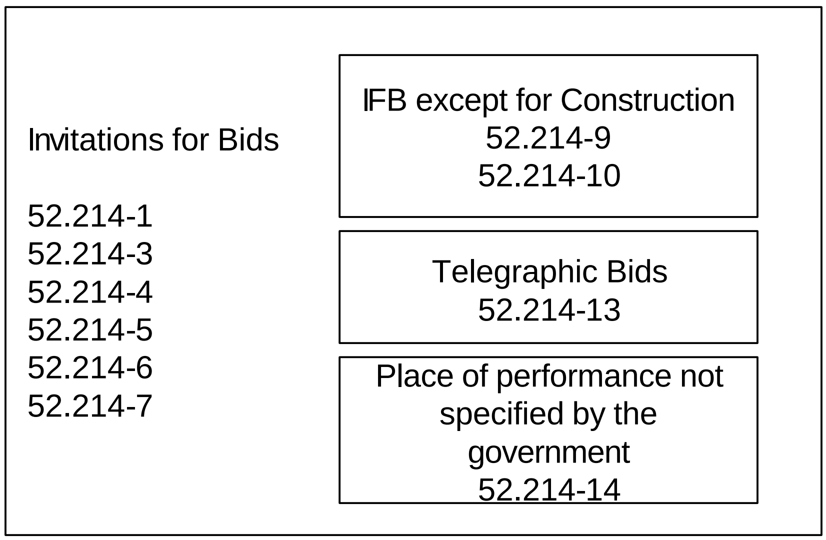

Figure 21. Provisions identified by 14.201-6

While it is possible that an IFB for construction could specify the place of performance (construction site) it is not a necessary condition. For example, the government might not require the construction of mobile shelters be done in a specific place and so whether the IFB is for construction and whether the

[87] 48 CFR 14.201-6(a).

government specifies a place of performance are independent factors. Similarly, the place of performance and whether the IFB is for construction are independent of whether the CO authorizes telegraphic bids.

The rules for incorporating the above provisions into instruments are:

- If IFB then insert 52.214-1, 52.214-3, 52.214-4, 52.214-5, 52.214-6, 52.214-7.
- IF IFB and not construction then insert 52.214-9 and 52.214-10.
- If IFB and telegraphic bids authorized then insert 52.214-13.
- IF IFB and place of performance is not specified by the government then insert 52.214-14.

Table 49 illustrates the encoding of these four rules.

Table 49. Encoding Provisions in Figure 21

Provision	IFB	Construction	Telegraphic Bids authorized	Place of performance specified by the government
52.214-1	T	I	I	I
52.214-3	T	I	I	I
52.214-4	T	I	I	I
52.214-5	T	I	I	I
52.214-6	T	I	I	I
52.214-7	T	I	I	I
52.214-9	T	F	I	I
52.214-10	T	F	I	I
52.214-13	T	I	T	I
52.214-14	T	I	I	F

Additional conditions would be required to assess the inclusion or exclusion of the following additional provisions:

1. 52.214-22 Evaluation of Bids for Multiple Awards is inserted in IFBs if the contracting officer determines that multiple awards might be made as a result of the solicitation, which is preconditioned upon the CO determining that multiple awards might be economically advantageous to the government.
2. 52.214-31, Facsimile Bids is inserted in IFBs when facsimile bids are authorized.
3. 52.214-12, Preparation of Bids is inserted in IFBs when the uniform contract format applies.
4. 52.214-23 Late submissions, Modifications, Revisions and Withdrawals of Technical Proposals under Two-step Sealed Bidding is inserted in solicitations for technical proposals in step-one of two-step sealed bidding.
5. 52.214-24, Multiple Technical Proposals is inserted in solicitations for technical proposals in step one of two-step sealed bidding if the contracting officer permits the submission of multiple technical proposals.
6. 52.214-25, Step Two of Two-Step Sealed Bidding, is inserted in IFBs issued under step two of two-step sealed bidding.

The first three provisions each require an additional condition. The last three address an approach to the process of sealed bidding that requires solicitation of a technical proposal from prospective bidders prior to soliciting a bid.

Sealed bidding subsumes both a single step and a two-step solicitation process. It is possible that telegraphic bids may not be authorized in step one of a two-step process but be authorized in step two of the two-step process. This situation is not evident in the prescription for applicability and use of the provision 52.214-13, Telegraphic Bids.[88]

Deterministic inductive logic facilitates answering questions like: Is it possible for a solicitation process to both authorize and prohibit telegraphic bids? In the case of the two-step sealed bidding process, it is conceivable that a contracting officer may not be willing to accept a telegraphic technical proposal while at the same time authorizing bidders to telegraph price information under step-two of two-step sealed bidding.

The ten features required to address the sixteen provisions discussed above are listed in Table 51.

The encoding for the sixteen provisions is illustrated in Table 52. Possibilities are encoded with the truth-value *indeterminate*.

---

[88]The applicable prescriptive paragraph is 48 CFR 14.201-6(g) "(1) Insert the provision at 52.214-13, Telegraphic Bids, in invitations for bids if the contracting officer decides to authorize telegraphic bids."

Table 50. Encoding of Additional Provisions at 14.201-6

Provision	IFB	Multiple Award anticipated	Facsimile Bids authorized	UCF	Step one of two step	Step two of two step	Mult tech props auth
52.214-22	T	T	I	I	I	I	I
52.214-31	T	I	T	I	I	I	I
52.214-12	T	I	I	T	I	I	I
52.214-23	T	I	I	I	T	I	I
52.214-24	T	I	I	I	T	F	T
52.214-25	T	I	I	I	I	T	F

The process of provision selection can be illustrated by specifying some conditions for a hypothetical acquisition. For the purpose of demonstration, we will assume that the instrument required is a solicitation in step two of two-step sealed bidding; that it is for construction; that telegraphic and facsimile bids are authorized; that the government does not specify the place of performance; and that the award will be made to a single bidder. We will further assume that during step-one, multiple technical proposals were authorized. These characteristics are illustrated in Table 53.

Table 51. Features Required by Provisions at 14.201-6

Feature	Condition
F(1)	Invitations for Bids
F(2)	Construction
F(3)	Telegraphic bids are authorized
F(4)	Place of performance specified by the government
F(5)	Multiple award anticipated
F(6)	Facsimile bids authorized
F(7)	Uniform contract format applicable
F(8)	Step one of two step sealed bidding
F(9)	Step two of two step sealed bidding
F(10)	Multiple technical proposals authorized

Table 52. Provision Encoding Using the Features Identified in Table 53

Provision	F(1)	F(2)	F(3)	F(4)	F(5)	F(6)	F(7)	F(8)	F(9)	F(10)
52.214-1	T	I	I	I	I	I	I	I	I	I
52.214-3	T	I	I	I	I	I	I	I	I	I
52.214-4	T	I	I	I	I	I	I	I	I	I
52.214-5	T	I	I	I	I	I	I	I	I	I
52.214-6	T	I	I	I	I	I	I	I	I	I
52.214-7	T	I	I	I	I	I	I	I	I	I
52.214-9	T	F	I	I	I	I	I	I	I	I
52.214-10	T	F	I	I	I	I	I	I	I	I
52.214-13	T	I	T	I	I	I	I	I	I	I
52.214-14	T	I	I	F	I	I	I	I	I	I
52.214-22	T	I	I	I	T	I	I	I	I	I
52.214-31	T	I	I	I	I	T	I	I	I	I
52.214-12	T	I	I	I	I	I	T	I	I	I
52.214-23	T	I	I	I	I	I	I	T	I	I
52.214-24	T	I	I	I	I	I	I	T	F	T
52.214-25	T	I	I	I	I	I	I	I	T	I

Table 53. Hypothetical Acquisition Conditions

Feature	Condition	Acquisition Condition
F(1)	Invitations for Bids	T
F(2)	Construction	T
F(3)	Telegraphic bids are authorized	T
F(4)	Place of performance specified by the government	F
F(5)	Multiple award anticipated	F
F(6)	Facsimile bids authorized	T
F(7)	Uniform contract format applicable	T
F(8)	Step one of two step sealed bidding	F
F(9)	Step two of two step sealed bidding	T
F(10)	Multiple technical proposals authorized	T

The provisions that are included and excluded in the hypothetical instrument are illustrated in Table 54. The conditions for inclusion are that the provision be in a class that is subsumed by the class defined by the instrument's conditions.

The sixteen provisions represent clear, uncomplicated situations. In 14.201-6(w)[89] the inclusion of 52.214-34, Submission of Offers in the English Language, requires the evaluation of a condition that is based alternatively upon whether any of the clauses prescribed in 25.1101 or 25.1102 are incorporated in the proforma contract which is anticipated to be awarded as a result of the

[89] 48 CFR 14.201-6(w) "Insert the provision at 52.214-34, Submission of Offers in the English Language, in solicitations that include any of the clauses prescribed in 25.1101 or 25.1102. It may be included in other solicitations when the contracting officer decides that it is necessary."

solicitation or upon the decision of the CO that the provision should be included. In 14.206-1(x)[90] the inclusion of 52.214-35, Submission of Offers in U.S. Currency, requires the evaluation of a condition that is based upon whether any of the clauses prescribed in 25.1101 or 25.1102 are incorporated in the proforma contract which is anticipated to be awarded as a result of the solicitation; unless the CO includes the clause at 52.225-17, Evaluation of Foreign Currency Offers as prescribed in 25.103(d) in that proforma contract. In both cases there is a dependency between the provisions selected for inclusion in the solicitation instrument and the contract clauses that have been selected as necessary or appropriate for the management of the acquisition post-award.

A condition like "in solicitations that include any of the clauses prescribed in 25.1101 or 25.1102" can be implemented in multiple ways. One approach is to construct a derived feature that can be used to encode "clauses prescribed in 25.1101 or 25.1102" and the other approach is to define features for each of the clauses incorporated in those sections. If the implementation were to define a derived feature and infer the value of that feature by evaluating a deductive disjunction of the clauses identified by those sections; if that clause list were to

[90] 48 CFR 14.201-6 (x) "Insert the provision at 52.214-35, Submission of Offers in U.S. Currency, in solicitations that include any of the clauses prescribed in 25.1101 or 25.1102, unless the contracting officer includes the clause at 52.225-17, Evaluation of Foreign Currency Offers, as prescribed in 25.1103(d). It may be included in other solicitations when the contracting officer decides that it is necessary."

Table 54. Included and Excluded Provisions Given the Hypothetical Acquisition Conditions in Table 53

	F(1)	F(2)	F(3)	F(4)	F(5)	F(6)	F(7)	F(8)	F(9)	F(10)
Instrument	**T**	**T**	**T**	**F**	**F**	**T**	**T**	**F**	**T**	**T**
Include										
52.214-1	T	I	I	I	I	I	I	I	I	I
52.214-3	T	I	I	I	I	I	I	I	I	I
52.214-4	T	I	I	I	I	I	I	I	I	I
52.214-5	T	I	I	I	I	I	I	I	I	I
52.214-6	T	I	I	I	I	I	I	I	I	I
52.214-7	T	I	I	I	I	I	I	I	I	I
52.214-13	T	I	T	I	I	I	I	I	I	I
52.214-14	T	I	I	F	I	I	I	I	I	I
52.214-31	T	I	I	I	I	T	I	I	I	I
52.214-12	T	I	I	I	I	I	T	I	I	I
52.214-25	T	I	I	I	I	I	I	I	T	I
Exclude										
52.214-9	T	F	I	I	I	I	I	I	I	I
52.214-10	T	F	I	I	I	I	I	I	I	I
52.214-22	T	I	I	I	T	I	I	I	I	I
52.214-23	T	I	I	I	I	I	I	T	I	I
52.214-24	T	I	I	I	I	I	I	T	F	T

change, the change could be implemented in the location where the feature is derived. If the implementer were to enumerate the applicable clauses as individual features, the process of constructing a feature vector to encode the applicability and use of these two provisions would require referring to the named sections. In very complex situations, such as those presented by the implementation of the FAR, reducing the number of dependencies that must be reviewed during a maintenance update may improve the correctness of the system and reduce the time and cost for maintenance.

Defining feature vectors to represent contracting officer preferences authorized in the acquisition context

The treatment of CO options can be illustrated from the terms for the applicability and use of the Value Engineering clause in supply or service contracts. Value engineering is essentially a program intended to provide incentives to contractors to identify and propose changes that could result in cost savings to the government. For simplification of the illustration we will address only some of the conditions that control the inclusion of the one of the possible value engineering clauses in a contract.

In a delegation of authority, any intervening authority may render the decision deterministic. For example, a value engineering program may not be cost effective for the types of supply or service contracts that are handled by the FLETC that would come within the provisions of 48.201(a) for which the clause might be optional (actions with a contract value of less than $100,000) and so the head of the contracting office at FLETC might preclude the option.

48.201 Clauses for supply or service contracts.

(a) *General*. The contracting officer shall insert a value engineering clause in solicitations and contracts when the contract amount is expected to be $100,000 or more, except as specified in subparagraphs (a)(1) through (5) and in paragraph (f) below. A value engineering clause may be included in contracts of lesser value if the contracting officer sees a potential for significant savings. Unless the chief of the contracting office authorizes its inclusion, the contracting officer shall not include a value engineering clause in solicitations and contracts--

(1) For research and development other than full-scale development;

(2) For engineering services from not-for-profit or nonprofit organizations;

(3) For personal services (see Subpart 37.1);

(4) Providing for product or component improvement, unless the value engineering incentive application is restricted to areas not covered by provisions for product or component improvement;

(5) For commercial products (see Part 11) that do not involve packaging specifications or other special requirements or specifications; or

(6) When the agency head has exempted the contract (or a class of contracts) from the requirements of this Part 48.

Figure 22. Applicability conditions for Value Engineering

In Table 55, features F(n) and F(n+1) are encoded to represent the applicable conditions and therefore unless the clause is precluded from use, those features should be encoded with truth-values of *true*. If a clause is explicitly excluded from use by the chief of the contracting office, then the features F(n) and F(n+1) should both be encoded with truth-values of *false*. The *false* truth-value in F(n+1) can be used to avoid asking the user whether they have a preference if their authority to make a choice has been preempted.

Table 55. Features for Controlling Optional Inclusion of Value Engineering

Feature identifier	Feature name	Applicability and use
Feature (n)	*Value-Engineering-clause-election-permission*[91]	*true*— chief of the contracting office authorizes its inclusion *false*— chief of the contracting office does not authorize its inclusion
Feature (n+1)	*Value-Engineering-clause-election*	*true*— contracting officer sees a potential for significant savings *false*— contracting officer does not see a potential for significant savings

One approach to setting default values for optional conditions is to assign *true* to the value of the feature that is used to describe the *acquisition context* and *true* to the value of the feature that is used to define the *acquisition action context*. This approach to default coding is interpreted to indicate that an option that may be elected, has been selected, pending de-selection by the CO. The result of this approach is to include any provision or clause that might be appropriate. Other

[91] The illustration has been simplified. In a realistic implementation a number of features may be required to account for specific cases in which the chief of the contracting office has authorized or not authorized inclusion of a Value Engineering clause in specific acquisition situations.

conditions may exclude a provision or clause on some other basis, however, this default treatment would have no effect on selection/exclusion unless the acquisition action context is a subset of the acquisition context defined for that provision or clause.

Table 56. CO Option Encoding

	F(n)	F(n+1)
52.248-1, Value Engineering	T	T
Contract Action	T	T

In the contract action feature vector the preference of the chief of the contracting office should be inherited from the clause encoding. If the clause is encoded with a truth-value *false* the clause should be excluded. If the default value for a CO's preference in an optional situation is encoded with a truth-value of *unknown* it will preclude automatic insertion of the clause and signal that a value is required to resolve applicability. Similarly, if the feature F(n+1) had been encoded for the clause with the truth-value *true* and defaulted for the action to *unknown*, the presence of the value *unknown* could be used to signal that a preference may be selected by the CO.

The situations just described can be reduced to four different situations illustrated in Tables 57-60.

Table 57. Default Authority - Default Contract Action

Default Authority Default Contract Action	F(n)	F(n+1)
52.248-1, Value Engineering	T	T
Contract Action (Case 1)	T	U

Table 58. Superior Authority Exclusion - Default Contract Action

Superior Authority Exclusion Default Contract Action	F(n)	F(n+1)
52.248-1, Value Engineering	F	F
Contract Action (Case 2)	F	U

Table 59. Superior Authorizes Inclusion - Explicit Exclusion by the CO

Superior Authorizes Inclusion Explicit exclusion by the CO	F(n)	F(n+1)
52.248-1, Value Engineering	T	T
Contract Action (Case 3)	T	F

In Table 57 the clause is not included until or unless the contracting officer revises F(n+1) to reflect a preference to include the clause. In this situation, the system should be designed to request a value from the user to replace the *unknown* value, however, if the user does not supply a value the system will nevertheless operate correctly.

Table 60. Superior Authorizes Inclusion - CO Elects Inclusion

Superior Authorizes Inclusion Contracting Officer Election	F(n)	F(n+1)
52.248-1, Value Engineering	T	T
Contract Action (Case 4)	T	T

In Table 58 the decision to preclude use by a higher authority results in the clause being explicitly excluded. In this situation, the system should be designed not to request a value from the user to replace the *unknown* value, since the user's preference is not relevant.

In Table 59 the CO has been delegated authority and has elected to exclude the clause. In Table 60 the CO has been delegated authority and has elected to include the clause.

The treatment of authority demonstrates that contingent decisions add complexity to the system design. Where a number of conditions are all contingent upon each other, a derived feature may be defined and its value inferred from the appropriate individual feature truth-values. This approach can be illustrated using the Value Engineering example in Table 61 when a clause is excluded by feature F(n) and included by feature F(n+1); and where F(n) has precedence over F(n+1); a new feature F(n+2) may be defined to resolve the potential conflict.

Table 61. Precedence Relationships

Superior Authority Permission Contracting Officer Election	F(n)	F(n+1)	Correct result	If(F(n) or F(n+1) = *false* then F(n+2) = *false* else F(n+2) =Combine(F(n), F(n+1))
Contract Action (Case 1)	T	U	Ask CO	U
Contract Action (Case 2)	F	U	Exclude	F
Contract Action (Case 3)	T	F	Exclude	F
Contract Action (Case 4)	T	T	Include	T

To evaluate a precedence relationship in which F(n) has precedence over F(n+1) and given the assignment rules described above, the value of F(n+2) can be determined with a conditional of the form:

IF(F(n) or F(n+1)) = *false*

THEN F(n+2) = *false*

ELSE F(n+2) = Combine(F(n), F(n+1)).

The interpretation of the truth-values of F(n+2) determined by that conditional should be to:

1. exclude the clause when F(n+2) has a truth-value of *false*
2. ask the CO if they have a preference when the truth-value is *unknown* and
3. include the clause when the truth-value is *true*.

The conditional expression should be used in cases where the goal of the inferred feature is to express preferential exclusion.[92]

[92] Preferential exclusion is the simplest approach to modeling a system in which authority is delegated, since if authority is not delegated, a TV(2) value can block any lower tier authority from acting independently since all lower tier authorities are limited by the authority delegated.

Chapter VII – Relationships between deterministic inductive logic and deductive logic

Deterministic inductive logic can be used, as a complement to deductive logic or as an alternative approach to representing propositions. Table 62 illustrates the mapping between Boolean expressions and their equivalent deterministic inductive logic feature vector representations.[93]

With deterministic inductive logic, any set of propositions can be combined by using the COMBINE( ) operator. The result of combinations takes into account quantification, where *true* and *false* are indicative of universal quantification and *indeterminate* is an indication of existential quantification. Tables 63, 64, and 65 illustrate the mapping of a few example combinations in both Boolean and feature vector forms.

In Table 63, the discriminating proposition was distinct from the propositions provided, where the two propositions make contradictory claims on the appropriateness of the descriptor b.

---

[93] Each Boolean formulation is associated with a line number (e.g., L4). These line numbers are used throughout this chapter as a shorthand notation for the original Boolean expression illustrated in this table.

Table 62. Mapping Boolean Expressions to Deterministic Inductive Logic

Line number	Boolean formulation		a	b	c
L1	a		T		
L2	not(a)		F		
L3	b			T	
L4	not(b)			F	
L5	c				T
L6	not(c)				F
L7	a and b		T	T	
L8	not(a and b)		I	I	
L9	not(a) and b		F	T	
L10	a and not(b)		T	F	
L11	a and c		T		T
L12	not(a and c)		I		I
L13	not(a) and c		F		T
L14	a and not(c)		T		F
L15	a and b and c		T	T	T
L16	not(a and b and c)		I	I	I
L17	a and not(b and c)		T	I	I
L18	not(a) and b and c		F	T	T
L19	not(a and c) and b		I	T	I
L20	(a and c) and not(b)		T	F	T
L21	(a or b)		I	I	
L22	(a or b or c)		I	I	I

Table 64 illustrates a situation in which one proposition subsumes the second.

Table 63. Mapping Compound Expressions – Contradictions

Array representation	Boolean formulation	A	b	C	Comments
L4	not(b)		F		
L9	not(a) and b	F	T		
COMBINE(L4, L9)	not(b) and not(a) and b	F	I		L2 (not(a)) is the discriminating proposition

Table 64. Mapping Compound Expressions – Subsumption

Array representation	Boolean formulation	A	b	C	Comments
L3	b		T		
L19	not(a and c) and b	I	T	I	
COMBINE(L3, L19)		I	T	I	L19 (b and not(a and c)) is the discriminating proposition

Table 65. Mapping Compound Expressions – Incomplete Representation

Array representation	Boolean formulation	a	b	C	Comments
L16	not(a and b and c)	I	I	I	
L17	a and not(b and c)	T	I	I	
L18	not(a) and b and c	F	T	T	
COMBINE(L16, L17, L18)		I	I	I	There is no discriminating proposition

Table 65 illustrates a situation in which the representation is not sufficient to discriminate the result set from the collection. Table 63 demonstrates a contradiction (b and NOT(b)). If this statement to be *true* (p = ~p), then *true* = *false* and *false* = *true*. The *indeterminate* truth-value handles this situation without the need to identify contradictions.

Set Interactions

A general diagram of the interactions among three sets {A, B, C} can be used to demonstrate set interactions. The three sets interact in seven possible regions as shown in Figure 23.

The diagram has a box surrounding the intersecting circles. The region not contained in the three circles that is within the box is implicitly asserted by the formulation NOT(A OR B OR C). This region contains two representations {8, 9} where 8 is intended to denote a new category defined as *not elsewhere classified* and "9" is retained to represent the *unknown* or residual present when negating all *known* categories.

The category *not elsewhere classified* can be added to any set of categories defined to distinguish between sub-facets of a category. This special class insures that the set of sub-facets fulfills the constraint that for facets must be collectively exhaustive.

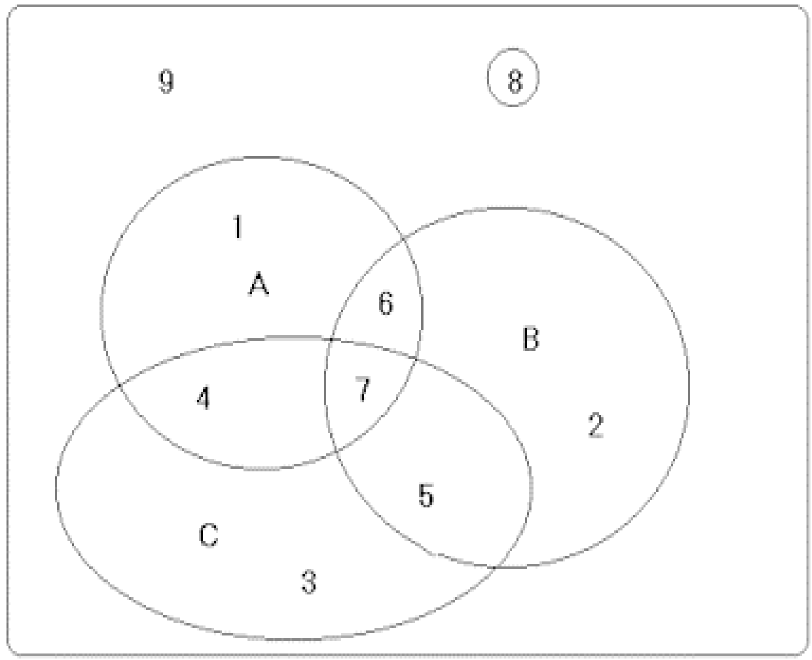

Figure 23. Regions defined by the interactions of three sets

There are three alternative approaches for handling feature vectors defined by enumerated sets.

In the first approach, a new feature may be created and assigned a state value on the basis of enumeration where the feature has value *true* for members of the enumerated set and *false* in every object other than the enumerated set. The first alternative addresses capturing membership. This approach assigns the truth-value *true* to the feature that is used to record set membership if the object is known to be a member of a set and *false* if the object's membership is not known.

Table 66. Set Interactions in Figure 23 Mapped to Boolean Expressions

Boolean formulation	Region numbers
A	1, 4, 6, 7
B	2, 5, 6, 7
C	3, 4, 5, 7
A AND B	6, 7
A AND C	4, 7
B AND C	5, 7
A AND B AND C	7
A AND NOT(B OR C)	1
B AND NOT (A OR C)	2
C AND NOT(A OR B)	3
NOT (A or B or C)	8 = not elsewhere classified
NOT(A or B or C)	9

In the second approach, a new feature may be created to account for all the not elsewhere classified cases and assigned a truth-value of *false* for members of the enumerated set and *unknown* in every object other than the enumerated set. This approach assumes that all objects that are not members of the set must belong to the not elsewhere classified set. Objects that are not members are assigned the value *true* and objects that are members are assigned the value *false*. This approach is the reverse of the first.

In the third approach, a new feature may be hypothesized as the discriminant category of the enumerated set, and assigned a truth-value of *true* for set members and a truth-value of *false* in every object not in the enumerated set. The third approach is a restatement of the first, however, it is based upon a presumption that whatever characteristic discriminates an object's membership, even though it may not be recognizable or even determinable directly, is *true* for ever member and *false* for every non member.

In each of the three approaches, assignments are inferred from assertions about the membership or non-membership of an object. These are alternative approaches for implementing systems when truth-value assignments are made on the basis of a rule like “I know it when I see it.”

Table 67 illustrates the three treatments using feature vectors with category 8 representing the first treatment, 9 the second, and 10 the third.

Completeness

Completeness allows truth-values to be applied to categories by inference. For example in Table 67, in L’1, if the item is correctly represented as “A” then the categories which can assume truth-value *true* given “A” {1, 4, 6, 7} can be enumerated exhaustively. Similarly, the categories which assume truth-value *false* given “A” {2, 3, 5, 8, 9} can be enumerated exhaustively. If, however, this system were not complete with respect to the number of categories and relationships between and among categories, it would not be possible to assign *false* truth-values implicitly from the proposition “A.”

Table 67. Approaches for Handling Enumerated Sets

Line referenc	**Boolean form**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
L'1	A	T	F	F	T	F	T	T	F	F	T
L'2	B	F	T	F	F	T	T	T	F	F	T
L'3	C	F	F	T	T	T	F	T	F	F	T
L'4	D	F	F	F	F	F	F	F	T	U	F
L'5	A and B	F	F	F	F	F	T	T	F	F	T
L'6	A and C	F	F	F	T	F	F	T	F	F	T
L'7	B and C	F	F	F	F	T	F	T	F	F	T
L'8	A and B and C	F	F	F	F	F	F	T	F	F	T
L'9	not (A)	F	T	T	T	T	F	F	T	U	T
L'10	not (B)	T	F	T	F	F	F	F	T	U	T
L'11	not (C)	T	T	F	F	F	T	F	T	U	T
L'12	not (D)	T	T	T	T	T	T	T	F	U	T
L'13	not (A and B and C and D)	F	F	F	F	F	F	F	F	U	F
L'14	A and not(D)	T	F	F	T	F	T	T	F	F	T

Proposition interactions

The examples L'1 through L'8 in Table 67, illustrate a number of different ways categories can interact in conjoint sets to define *inclusionary rules*. Examples L'9 through L'13 demonstrate *exclusionary rules*. Example L'14 illustrates a potentially problematic type of inclusionary and exclusionary rule combination. The L'14 case demonstrates disjoint sets where one specifies an inclusionary proposition and the second an exclusionary proposition. In L'14, "A and NOT(D)" is pragmatically equivalent to L'1, which is the more restrictive of the two propositions. Disjoint sets used in combination in which one is inclusionary and the other exclusionary can be treated using a simple law of set interaction. If two propositions are specified in which, one 'a' is used as an inclusionary rule and the other 'b' as an exclusionary rule, and the conjoint set defined by combining the propositions 'a AND b' is equal to the null or empty set, then the most restrictive of the two propositions may be used independently to define a minimal result set.

The discriminatory power of exclusionary rules of the types in L'9 through L'13 of Table 67, while presenting the potential for problems of the type represented in L'14, can be used for describing situations in which a user rejects examples as "not relevant." Exclusionary rules should be formulated and employed whenever they contribute to efficiently discriminating a desired result set.

Chapter VIII Discussion, conclusions and future research

Discussion

DIL is intended to be descriptive and not predictive. The result of this research has been to add an inductive operator to complement first order deductive logic. DIL is intended to support combinations of descriptive feature vectors to identify universally quantified features. Universally quantified features are logically necessary conditions for membership in a set defined by the combination operator.

DIL only addresses the concept of sufficiency in terms of the whether there are any universally quantified features that distinguish a class of interest from other classes. There may be a number of additional features that are not logically necessary features that may be required to distinguish between objects with different degrees of relevance depending upon the level of abstraction of a specific class. In this sense, all the necessary conditions are insufficient to distinguish relevant documents. Whenever the set of necessary features is less than the number of sufficient features there is a mismatch between the representation of objects and the representation required to fully and completely represent a specific user need.

In situations in which the feature set available is insufficient to accurately and completely represent a user's information needs, a filtering or retrieval system must operate with less than perfect performance. In these situations the best alternative system is a function of the pragmatic performance of alternative systems relative to the specific information needs under analysis. Any environment in which there is an inherent degree of error in filtering and retrieval results is pragmatically a probabilistic environment. In a probabilistic environment, a controlled vocabulary approach may not be useful. In such situations a range of other inductive information retrieval and filtering techniques that build models of user's needs inductively from relevant examples (training data) would be more appropriate, productive and cost effective than the application of DIL. If DIL could be applied as an aid for managing user decisions about object categories, it would require integration with an automated retrieval or filtering system. The role of DIL in such a situation would be solely to build a hierarchy of concepts and potentially to associate the user defined conceptual categories with tools to facilitate access and relevance feedback processing. This type of application would be more productively accomplished with a system of the type described by Michalski.[94]

It is conceivable that a feature in a DIL feature vector could be associated with a statistical classifier in an implementation in which that classifier can provide the required level of performance. To implement DIL would require creating object surrogates to store DIL feature values and the necessary implementation support to create and maintain feature vector values in

---

[94] Michalski and Stepp, *Learning from Observation.*

conformance with which objects that are selected by which classifiers. Integrating DIL with quantitatively based information filtering or retrieval techniques is beyond the scope of this research, however, it is more than likely essential to constructively integrating DIL into any commercial information systems.

Novelty claims

The novel contribution of this research is the definition of a deterministic, inductive combination function. The contribution is an inductive conjunctive operator (e.g., COMBINE()), which may provide a variety of potential benefits in a number of different situations. The benefits of the COMBINE() operator in the design of algorithms and systems is subject to empirical observation and future research.

There is nothing novel or original in the definition of either the compare or contrast functions. Both the compare and contrast functions implement nothing more than a test for equality or inequality that has been adapted for a multi-valued logic. The compare and contrast functions are essentially character or string comparison functions that return a binary logic vector with a length equal to the length of the strings being compared.

DIL uses feature vectors. Feature vectors have been used in many applications. There is nothing new or novel about a feature vector. The feature vectors implemented in DIL are qualitative feature vectors. The values of features in DIL are nominal (categorical) and may be structured (multiple valued) if the feature contains categories that are mutually exclusive, collectively exhaustive and number more than two. A DIL feature vector is a sequence of truth-values in the same way a Horn clause is a sequence of truth-values. While the number of truth-values of DIL features is different from that of Horn clauses, this is not a sufficient basis for any claim of novelty to an ordered sequence of truth-values.

DIL is a multi-valued logic. There are a number of multi-valued logics. A number of different logic theories based upon multiple values of truth have been formulated, such as those of Lukasiewicz, Bochvar, Kleene, Heyting, and Reichenbach. Some common types are those based upon three values of truth representing *true*, *false* and *unknown*. These trivalent or three-valued logics commonly represent the three truth-values of *true*, *false,* and *unknown* by 1, 0, and 1/2 respectively. For example, Lukasiewicz developed an N-valued logic in the 1930's. In his N-valued logic, the set T_N of truth-values are assumed evenly divided over the closed interval {0,1}.[95]

DIL is based upon a mapping of probabilistic induction to the quantification types defined in First Order logic. DIL is conceptually a subset of probabilistic induction in which the linear scale of probabilistic induction has been mapped to three discrete states (T, F, I). A particular mapping of truth-

---

[95] Joseph Giarratano and Gary Riley, *Expert Systems: Principles and Programming,* (Boston:PWSKENT PublishingCompany, 1989) 318-321.

values is not a sufficient basis for asserting any novelty to the multi-valued logic aspects of DIL.

Implementation claims

DIL is not a system. It is a proposal for an approach to building qualitative feature vectors and a specification of a limited set of functions and relationships that might be useful in organizing, analyzing and managing information.

DIL is not an algorithm in that it has no inherent behavior. The three functions return values based upon arguments. A logical operator with two arguments (i.e., COMBINE(a,b)) is not conceptually an algorithm any more than a Boolean AND operator with two arguments (a and b) is an algorithm.

DIL is not designed to be a predictive model. We propose that one might specify filters in DIL in terms of inclusionary and exclusionary criteria, however, this type of filter is not at all the type of predictive information filter described by Oard.[96] A filter implemented in DIL is a logic based pattern match of the type typical of controlled vocabulary, exact match systems.

Research into machine learning, automatic indexing, information filtering, clustering, modern information retrieval, computational linguistics, text understanding, and natural language processing all use inductive processes to build generalizations from specifics. DIL has made no new or novel contribution

---

[96] Oard, *The State of the Art in Text Filtering*.

in any way to learning from examples. In fact, DIL is not a learning algorithm nor is it proposed as such. DIL is not a substitute for any machine-learning technologies or algorithms. While one might map the results of a learning algorithm into DIL, DIL has no inherent learning capabilities.

DIL is not a classifier in the way that term is used in computer science. DIL is not intended to provide the capabilities of a statistical classifier (an algorithm that develops a method for recognizing objects or events of a particular type from training data). Automated classification research has resulted in the development of computer-based tools for recognizing and classifying objects or events using probabilistic induction, genetic algorithms and neural networks. These approaches involve the use of training data sets to construct classifiers that, when implemented, can be used to assess events and objects and associate them with a type. Classifiers have been applied to a large variety of applications as diverse as information filtering and retrieval, natural language processing and the strategic defense initiative for identifying re-entry vehicles. DIL can be used in association or conjunction with algorithms used for classifying events and objects, however, it cannot contribute in any way to accomplishing the objective of automated classification or indexing.

Multi-faceted classification example

In Chapter three a multi-faceted model was constructed using DNA as the basis of the illustration. The intent of the illustration was not to make any claims about the value of DIL in organizing, managing or analyzing genetic information, but only to illustrate a faceted model. If DIL can be productively applied to the analysis and management of genetic information, its limits of usefulness are in identifying the homogeneous characteristics of sets of data. It is not currently feasible to construct feature vectors that represent the 2.4 million single nucleotide polymorphisms (SNPs) identified by Celera that are expressed in DNA samples collected from individuals. In fact, at the present state of research, it is not clear how widely accepted the estimate of the number of SNPs is. The magnitude estimate does provide a scale factor that may be useful in conceptualizing DIL's relative computational efficiency for its limited usefulness in identifying the homogeneous characteristics of groups of samples. The computational effort required to process 100 cases of 2.5 million features using DIL's approach is less than other approaches for identifying the characteristics of homogeneity.

Identifying the characteristics of homogeneity has no effect on either predicting disease or on identifying precursors for disease. In fact, the homogeneous characteristics of a particular population provide little or no information that is useful in modeling causes or effects in a population. The sole benefit of a quick, low computational cost tool, such as DIL, is to provide a basis for searching for additional individuals (i.e., increase recall) that are not in the known population. Identifying additional individuals that express all the characteristics shared by a known group may contribute additional information useful in solving a problem. New members might lead to subdividing the original population or to some other conclusion entirely.

DIL provides only a single new logical operator that can be used in manipulating genetic information. The usefulness, if any, is a function of how the operator is applied and used by researchers.

Natural language applications

DIL is not, in and of itself, useful to natural language processing, text understanding or natural language understanding processes or systems. DIL may be useful in managing feature sets or identifying relevant features in natural language applications only if appropriate tools used for natural language analysis assign feature values and when and if those features can be usefully linked to some aspect of interest to a user. For example, a program designed to parse and analyze text might produce a conceptual graph, where a mapping exists between the nodes and arcs of the graphs and DIL compatible feature vectors in a system that creates and maintains DIL surrogates that are mapped from the conceptual graphs. Where features are created and assigned, the combine function might assist in identifying relevant aspects (subjects, verbs, objects, referents) that are universally present or absent in sets of cases. Alone, DIL cannot make any contribution to parsing, assigning parts of speech, or doing any linguistic analysis.

Structured information

DIL relies on structured information and cannot operate with unstructured information such as text without external assistance in the form of a human programmer or in conjunction with other automation. DIL was designed for managing feature sets, where features are used to represent the presence or absence of specific characteristics. A parser that can identify lexical elements (the presence of words based upon sequences of letters) can be used to assign a feature on the basis of the presence of a specific term. For example, a lexical analyzer could be used to identify the term “walk” in a text. An approach for identifying the usage of a term, such as an augmented transition network, may be used to distinguish a term’s usage, for example in terms of parts of speech. Natural language processing supports adding information such as the function of a term in the communication. For example, if one were searching for travel related information it might be useful to distinguish objects that describe something akin to “a historical walk through the city of Charleston” in categorizing leisure activities. DIL can tract the assignment of a feature to record that an object contains information about a “leisure activity,” however, it is totally reliant upon external information to assign the feature value for any object.

Chapter six illustrated an application of DIL, which is essentially a rule based expert system type application. The objective of the FAR document application is to construct rules in the form of feature vectors that can be used to categorize texts and to select appropriate texts in specific circumstances. The example is typical of developing production rules in an expert system type application in which a set of rules is defined to trigger a particular outcome. A similar requirement is in systems used to record business rules in management information systems. This type of application is one in which a deterministically correct outcome is essential to the system objective. A deterministically consistent outcome is frequently the objective of an automated system in which either there is some potential economic liability associated with failures or there is some other form of criticality associated with the system's operation. When a system is sensitive to the presence or probability of errors, production rules may be an appropriate approach to design so that an acceptable error threshold can be achieved deterministically.

As previously stated, DIL is not capable of extracting the information necessary to construct feature vectors from the text of the FAR. The feature vectors in chapter six were developed by hand for illustrative purposes only. If the application were developed and deployed commercially, a programmer or knowledge engineer would be responsible for mapping the text of the FAR's prescriptions for applicability and use of standard texts to features and feature vectors.

Conceptual clustering

Conceptual clustering[97] provides a systematic approach for building classifications that are conceptually navigable. That research used (1) a set of physical or conceptual objects, represented in terms of a set of attributes, and (2) a body of background knowledge about problem constraints and defined an

---

[97] Michalski and Stepp, *Learning from Observation*.

algorithm (implemented as a computer program named CLUSTER/2) that was demonstrated to be capable of constructing conceptual classifications inductively. There were two distinct components, a clustering module and a module for building a hierarchy. The construction of clusters is an inductive process intended to form a set of disjoint sets of objects; while the assignment of those clusters into a hierarchy is done deductively by division.

There are some conceptual overlaps between the treatment of criteria for clustering quality described by Michalski and the specifications of DIL. Michalski's *inter-cluster difference* is a measure of the degrees of difference between a pair of complexes after removing selectors that intersect.[98] Using the automobile example from Chapter 3, a pair of complexes of the type described by Michalski might be:

- [color = black, *red*] [type = *CRV* or Passport]
- [color = blue, *red*] [type = *CRV* or Odyssey]

These complexes have a *degree of disjointness* of 4, because 2 of the 6 selectors intersect (intersecting selectors are italicized).[99]

In DIL a selector would be represented as a feature. DIL would describe a complex in terms of a feature vector and would determine the similarity or difference using the compare and contrast functions.

---

[98] Ibid., 345.

[99] Ibid.

Table 68. DIL Feature Vector

	black	red	blue	CRV	Passport	Odyssey
Complex(1)	T	T	F	T	T	F
Complex(2)	F	T	T	T	F	T
Contrast(1,2)	T	F	T	F	T	T

The calculation of DOD for the two vectors in Table 68 is analogous to the *inter-cluster difference* described by Michalski. In DIL the DOD for the data in Table 68 would return a value of 0.666 (4/6) which is a conceptually similar measure to that calculated by the inter-cluster difference formula. The difference between measures is that DOD is normalized to represent a scaled value (percent) as opposed to being an absolute measure of the number of differences between complexes. There is no inherent or material advantage to DOD over the inter-cluster difference measure. DIL provides a DOS measure which is a measure of similarity. A similarity measure is not an appropriate measure of the quality of clustering, but is rather a basis for associating an event with an existing cluster.

The *inter-cluster difference* measure is a metric that is used to assess the quality of clustering and represents an important operational component of the process. The DIL measures of DOD and DOS are simply descriptive and their usefulness, beyond equivalence testing, is dependent upon their interpretation and use by a user.

DIL processes information of the type used by Michalski in formulating selectors. Michalski refers to selectors that differentiate a conceptual category as logical-complexes and set-complexes. They are used to formulate the set of disjoint clusters that are later associated with the list of predefined conceptual categories that represent leaf nodes in a hierarchical classification. The hierarchical assignment begins with the universe represented by the entire collection and then divides it into successive subcategories. Clearly clustering is an inductive process and clearly the classification hierarchy that results from CLUSTER/2 is conceptually navigable, however, the formation of the classification is totally dependent upon an *a priori* conceptual scheme and hence is not substantially similar to DIL.

Conclusions

As a result of defining deterministic inductive logic, there is now a deterministic inductive complement to classical deductive logic and a deterministic form of inductive logic that is an alternative to probabilistic induction.

DIL is not proposed as a substitute for existing information management tools, but as an alternative that may be useful in some situations. For example, research into information filtering and retrieval using quantitative methods has resulted in a number of systems approaches that are universally applicable at some level of performance as measured by recall, precision, discrimination and fallout. DIL is not a substitute for any of the quantitative methods that have been effectively applied to information filtering and retrieval. However, as the volume of information in systems continually increases, a capability to reduce the size of solution sets that result from information filtering or retrieval is essential to avoiding information overload.

To effectively remove non-relevant information requires a way to use relevance feedback in ways that can exclude records that are not of interest to a user. Exclusionary rules are inherently dangerous in that there is a potential for unintended consequences when exclusionary rules exclude relevant information. Discriminants, as defined in DIL, are universally quantified characteristics and when applicable do not result in unintended consequences. When relevant and non-relevant feature vectors are combined, the DIL result will cause any features that occur in both relevant and non-relevant objects to become *indeterminate*. By eliminating *indeterminate* features from exclusionary rules, no unintended exclusions occur, rendering the approach potentially safe for solution set reduction.

Traditional deductive classifications embed a fixed perspective through the sequence of divisions used to organize information. An inductive classification approach provides the potential for multiple perspectives to be modeled. DIL contributes an approach to inductively building conceptual structures that result in classifications similar to traditional classifications but with the potential for representing multiple perspectives.

The performance of an information filtering or retrieval system is a function of the ability of the system and user to represent information needs.

When a system's representation does not match the requirements of a user, either because concepts important to the user are not represented, or the specificity of concepts represented by the system is different from the level of specificity required by the user, the system's retrieval or filtering performance relative to the needs of the user will be suboptimal. DIL facilitates a user in defining their own conceptual categories by example and by assisting in building a personal vocabulary that differentiates concepts the user wishes to differentiate and associates ones the user considers similar. Representational sufficiency provides a measure of the sufficiency of a system's vocabulary to represent those concepts required by a user to organize a set of information objects into their own conceptual model. When a system vocabulary is insufficient, either the vocabulary should be extended or the user must rely on an inexact result such as quantitative ranking.

Ultimately the usefulness or efficacy of DIL must be a function of how it is applied and whether or to what extent deterministic induction contributes to organizing, analyzing and managing information in ways that facilitate a user's access, use and understanding of information.

Future research

The research thus far into DIL has focused on defining operators and ways in which operators can be combined to provide capabilities for organizing and analyzing information. This research has been theoretical and hasn't addressed how DIL might be applied, how to communicate DIL to potential users, and the potential benefits, if any, of applying DIL in a variety of different implementation situations. Having defined and demonstrated DIL and its relation to deductive logic, future research is needed that is focused on extending and enhancing DIL and applying DIL to information management and to information systems design.

DIL provides a new tool for information organization and analysis. It is possible that DIL may contribute to a number of different application areas, not so much as an alternative but as an enhancement or additional tool for:

- Data mining
- Construction of index languages
- Qualitative research
- Information filtering
- Expert systems

Data mining has traditionally used probabilistic models of the relative importance of a number of factors in a particular outcome or the relative importance of a number of possible indicators in predicting a particular outcome. For example, data mining is used by organizations to model potential user preferences or to predict potential customers for products and services. Ordinarily data mining begins with some *a priori* assumptions about what relationships are important and provides probabilities that a certain target customer would be an appropriate target for a specific product or service.

DIL presents the potential for analyzing information to identify if there are any deterministic characteristics or discriminants that can be used to differentiate one category or class of objects or cases given some observation or target outcome. We expect that, by creating sets that have characteristics that are homogeneous for all set members, decision makers may more easily conceptualize the relationships found in databases such as the characteristics of target audiences.

Human lexicographers, on the basis of surveys of users' information needs and subject description requirements, have traditionally manually built and maintained indexing languages and thesauri. The hierarchical relationships represented in these languages are selected to provide mechanisms for navigating the indexing language and identifying preferred terms, synonymous terms, broader, narrower and related terms. The design of an indexing language results in a deductive classification in which a subject is organized by successive division. DIL provides capabilities for constructing classifications from the bottom up and testing classifications for overlap, subsumption, etc. It is possible that DIL could contribute to the productivity of a lexicographer in analyzing and maintaining an indexing language. Research into constructing tools to support lexicographers in the design, construction and maintenance of indexing languages and thesauri may contribute to building system capabilities that ordinary users might exploit in analyzing and organizing information in accordance with conceptual models of their own creation.

The constant comparison technique is used in qualitative research to identify relationships. The relationships found in a qualitative study may be used in formulating or verifying a typology, which may be used to communicate the relationships between and among objects and interactions observed in some study's context. Deterministic inductive logic provides a capability for building representations, formally testing the relationships between cases and classes of cases, and assessing the presence or absence of discriminants. The inductive capabilities of DIL might assist an auditor of a naturalistic study in testing the consistency and completeness of another researcher's coding of study observations. The application of DIL to facilitate the constant comparison technique or to provide a semi-automated mechanism for the verification and audit of qualitative research studies is of interest as an area for future research.

As demonstrated in chapter six, in the application of DIL to document preparation, DIL provides a capability for building deterministic information filters that can be used to select information from an information stream or retrieve information from a collection. The contribution of DIL is the identification of discriminants that can be used to select or reject information objects. Whenever a system's information representation is not perfectly matched with the language and interests of users the mismatch results in probabilistic predictions of relevance as typified by ranking systems. DIL may contribute to reducing the noise in a probabilistic ranking system when, by integrating DIL into both the retrieval/filtering and relevance feedback processing components of an information system, the retrieval sets can be reduced to exclude objects that are known to be not relevant. To determine if DIL can make any contribution and to assess its relative contribution, if any, we anticipate conducting research into integrating DIL with quantitatively based information retrieval and filtering technologies to test the relative effects of DIL on recall, precision, discrimination and fallout.

The document generator described in chapter six is basically an expert system that uses rules for the inclusion or exclusion of standard texts. The approach described encodes for each possible text the terms of its applicability and use, where the applicability and use rules describe the appropriateness of a text in a particular outcome or context. Instead of integrating all the system's logic into system level decision trees, each possible outcome is encoded in terms of its applicability and use conditions, where the entire set of possible outcomes is filtered for a desired set of conditions at the time the expert system is executed. In situations such as the document preparation problem, where continual changes result in periodic system modifications and maintenance, the ability to isolate the components of the system that have changed may simplify maintenance. If every time there is a change, the entire expert system does not require re-verification and/or re-validation, but only changed components require quality testing, DIL may be a useful tool for building and maintaining expert systems.[100] Determining the relative contribution of DIL to the construction and maintenance of expert systems is an area of potential interest for future research.

---

[100] Verification is that portion of the quality assurance process in which the appropriateness of a design to fulfilling some set of operational intentions for a system is evaluated. In the case of expert systems, verification addresses determining whether a rule or rules are consistent and can be expected to fulfill system objectives. Validation is that portion of the quality assurance process in which a system's performance is tested during execution to determine if the system behaves in operation as designed. In the case of expert systems, validation addresses determining whether the rules encoded in the system execute individually and in combination as intended.

Appendix A – Index to FAR Provisions and Clauses referenced in Chapter VI

Number	Citation (Text)	Clause Title	Date	Type (provision/clause)
1	52.202-1	Definitions (Oct 1995)	(Oct 1995)	clause
		Alternate I	(Apr 1984)	clause
2	52.214-1	Solicitation Definitions--Sealed Bidding.	(Jul 1987)	provision
	52.214-2	[Reserved]		
3	52.214-3	Amendments to Invitations for Bids.	(Dec 1989)	provision
4	52.214-4	False Statements in Bids.	(Apr 1984)	provision
5	52.214-5	Submission of Bids.	(Mar 1997)	provision
6	52.214-6	Explanation to Prospective Bidders.	(Apr 1984)	provision
7	52.214-7	Late Submissions, Modifications, and Withdrawals of Bids.	(Nov 1999)	provision
	52.214-8	[Reserved]		
8	52.214-9	Failure to Submit Bid.	(Jul 1995)	provision
9	52.214-10	Contract Award--Sealed Bidding.	(Jul 1990)	provision
	52.214-11	[Reserved]		
10	52.214-12	Preparation of Bids.	(Apr 1984)	provision
11	52.214-13	Telegraphic Bids.	(Apr 1984)	provision

12	52.214-14	Place of Performance--Sealed Bidding.	(Apr 1985)	provision
13	52.214-15	Period for Acceptance of Bids.	(Apr 1984)	provision
14	52.214-16	Minimum Bid Acceptance Period.	(Apr 1984)	provision
	52.214-17	[Reserved]		
15	52.214-18	Preparation of Bids--Construction.	(Apr 1984)	provision
16	52.214-19	Contract Award--Sealed Bidding-- Construction.	(Aug 1996)	provision
17	52.214-20	Bid Samples.	(Apr 1984)	provision
		Alternate I	(Apr 1984)	provision
		Alternate II	(Apr 1984)	provision
18	52.214-21	Descriptive Literature.	(Apr 1984)	provision
		Alternate I	(May 1999)	provision
19	52.214-22	Evaluation of Bids for Multiple Awards.	(Mar 1990)	provision
20	52.214-23	Late Submissions, Modifications, Revisions, and Withdrawals of Technical Proposals under Two-Step Sealed Bidding.	(Nov 1999)	provision
21	52.214-24	Multiple Technical Proposals.	(Apr 1984)	provision
22	52.214-25	Step Two of Two-Step Sealed Bidding.	(Apr 1985)	provision
23	52.214-26	Audit and Records--Sealed Bidding.	(Oct 1997)	clause

24	52.214-27	Price Reduction for Defective Cost or Pricing Data--Modifications--Sealed Bidding.	(Oct 1997)	clause
25	52.214-28	Subcontractor Cost or Pricing Data--Modifications--Sealed Bidding.	(Oct 1997)	clause
26	52.214-29	Order of Precedence--Sealed Bidding.	(Jan 1986)	clause
27	52.214-30	Annual Representations and Certifications--Sealed Bidding.	(Jan 1997)	provision
28	52.214-31	Facsimile Bids.	(Dec 1989)	provision
	52.214-32	[Reserved]		
	52.214-32	[Reserved]		
29	52.214-34	Submission of Offers in the English Language.	(Apr 1991)	provision
30	52.214-35	Submission of Offers in U.S. Currency.	(Apr 1991)	provision

Appendix B – FAR Provisions and Clauses referenced in Chapter VI

48 CFR 52.202-xx Contract clauses prescribed by 48 CFR 2.2

52.202-1 Definitions.

As prescribed in Subpart 2.2, insert the following clause:

2.201 Contract clause.

The contracting officer shall insert the clause at 52.2021, Definitions, in solicitations and contracts except when the contract is not expected to exceed the simplified acquisition threshold. If the contract is for personal services, construction, architect-engineer services, or dismantling, demolition, or removal of improvements, the contracting officer shall use the clause with its Alternate I. Additional definitions may be included, provided they are consistent with the clause and the FAR.

Definitions (Oct 1995)

(a) "Head of the agency" (also called "agency head") or "Secretary" means the Secretary (or Attorney General, Administrator, Governor, Chairperson, or other chief official, as appropriate) of the agency, including any deputy or assistant chief official of the agency; and the term "authorized representative" means any person, persons, or board (other than the Contracting Officer) authorized to act for the head of the agency or Secretary.

(b) "Commercial component" means any component that is a commercial item.

(c) "Commercial item" means--

(1) Any item, other than real property, that is of a type customarily used for nongovernmental purposes and that--

(i) Has been sold, leased, or licensed to the general public; or

(ii) Has been offered for sale, lease, or license to the general public;

(2) Any item that evolved from an item described in paragraph (c)(1) of this clause through advances in technology or performance and that is not yet available in the commercial marketplace, but will be available in the commercial marketplace in time to satisfy the delivery requirements under a Government solicitation;

(3) Any item that would satisfy a criterion expressed in paragraphs (c)(1) or (c)(2) of this clause, but for--

(i) Modifications of a type customarily available in the commercial marketplace; or

(ii) Minor modifications of a type not customarily available in the commercial marketplace made to meet Federal Government requirements. "Minor" modifications means modifications that do not significantly alter the nongovernmental function or essential physical characteristics of an item or component, or change the purpose of a process. Factors to be considered in determining whether a modification is minor include the value and size of the modification and the comparative value and size of the final product. Dollar values and percentages may be used as guideposts, but are not conclusive evidence that a modification is minor;

(4) Any combination of items meeting the requirements of paragraphs (c)(1), (2), (3), or (5) of this clause that are of a type customarily combined and sold in combination to the general public;

(5) Installation services, maintenance services, repair services, training services, and other services if such services are procured for support of an item referred to in paragraphs (c)(1), (2), (3), or (4) of this clause, and if the source of such services--

(i) Offers such services to the general public and the Federal Government contemporaneously and under similar terms and conditions; and

(ii) Offers to use the same work force for providing the Federal Government with such services as the source uses for providing such services to the general public;

(6) Services of a type offered and sold competitively in substantial quantities in the commercial marketplace based on established catalog or market prices for specific tasks performed under standard commercial terms and conditions. This does not include services that are sold based on hourly rates without an established catalog or market price for a specific service performed;

(7) Any item, combination of items, or service referred to in subparagraphs (c)(1) through (c)(6), notwithstanding the fact that the item,

combination of items, or service is transferred between or among separate divisions, subsidiaries, or affiliates of a Contractor; or

(8) A nondevelopmental item, if the procuring agency determines the item was developed exclusively at private expense and sold in substantial quantities, on a competitive basis, to multiple State and local Governments.

(d) "Component" means any item supplied to the Federal Government as part of an end item or of another component.

(e) "Nondevelopmental item" means--

(1) Any previously developed item of supply used exclusively for governmental purposes by a Federal agency, a State or local government, or a foreign government with which the United States has a mutual defense cooperation agreement;

(2) Any item described in paragraph (e)(1) of this definition that requires only minor modification or modifications of a type customarily available in the commercial marketplace in order to meet the requirements of the procuring department or agency; or

(3) Any item of supply being produced that does not meet the requirements of paragraph (e)(1) or (e)(2) solely because the item is not yet in use.

(f) "Contracting Officer" means a person with the authority to enter into, administer, and/or terminate contracts and make related determinations and findings. The term includes certain authorized representatives of the Contracting

Officer acting within the limits of their authority as delegated by the Contracting Officer.

(g) Except as otherwise provided in this contract, the term "subcontracts" includes, but is not limited to, purchase orders and changes and modifications to purchase orders under this contract.

(End of clause)

Alternate I (Apr 1984)**.** If the contract is for personal services; construction; architect-engineer services; or dismantling, demolition, or removal of improvements, delete paragraph (c) of the basic clause.

(End of Alternate clause)

48 CFR 52.214-xx Solicitation provisions prescribed by 48 CFR 14.201-6 and contract clauses prescribed by 48 CFR 14.201-7

52.214-1 Solicitation Definitions--Sealed Bidding.

As prescribed in 14.201-6(b)(1), insert the following provision:

14.201-6 Solicitation provisions.

(b) Insert in all invitations for bids the provisions at--

(1) 52.214-1, Solicitation Definitions--Sealed Bidding;

Solicitation Definitions--Sealed Bidding (Jul 1987)

"Government" means United States Government.

"Offer" means "bid" in sealed bidding.

"Solicitation" means an invitation for bids in sealed bidding.

(End of provision)

52.214-2 [Reserved]

52.214-3 Amendments to Invitations for Bids.

As prescribed in 14.201-6(b)(3), insert the following provision:

14.201-6 Solicitation provisions.

(b) Insert in all invitations for bids the provisions at--

(3) 52.214-3, Amendments to Invitations for Bids; and

Amendments to Invitations for Bids (Dec 1989)

(a) If this solicitation is amended, then all terms and conditions which are not modified remain unchanged.

(b) Bidders shall acknowledge receipt of any amendment to this solicitation (1) by signing and returning the amendment, (2) by identifying the amendment number and date in the space provided for this purpose on the form for submitting a bid, (3) by letter or telegram, or (4) by facsimile, if facsimile bids are authorized in the solicitation. The Government must receive the acknowledgment by the time and at the place specified for receipt of bids.

(End of provision)

52.214-4 False Statements in Bids.

As prescribed in 14.201-6(b)(4), insert the following provision in all invitations for bids:

14.201-6 Solicitation provisions.

(b) Insert in all invitations for bids the provisions at--

(4) 52.214-4, False Statements in Bids.

False Statements in Bids (Apr 1984)

Bidders must provide full, accurate, and complete information as required by this solicitation and its attachments. The penalty for making false statements in bids is prescribed in 18 U.S.C. 1001.

(End of provision)

52.214-5 Submission of Bids.

As prescribed in 14.201-6(c)(1), insert the following provision:

14.201-6 Solicitation provisions.

(c) Insert the following provisions in invitations for bids:

(1) 52.214-5, Submission of Bids.

Submission of Bids (Mar 1997)

(a) Bids and bid modifications shall be submitted in sealed envelopes or packages (unless submitted by electronic means)--

(1) Addressed to the office specified in the solicitation; and

(2) Showing the time and date specified for receipt, the solicitation number, and the name and address of the bidder.

(b) Bidders using commercial carrier services shall ensure that the bid is addressed and marked on the outermost envelope or wrapper as prescribed in subparagraphs (a)(1) and (2) of this provision when delivered to the office specified in the solicitation.

(c) Telegraphic bids will not be considered unless authorized by the solicitation; however, bids may be modified or withdrawn by written or telegraphic notice.

(d) Facsimile bids, modifications, or withdrawals, will not be considered unless authorized by the solicitation.

(e) Bids submitted by electronic commerce shall be considered only if the electronic commerce method was specifically stipulated or permitted by the solicitation.

(End of provision)

52.214-6 Explanation to Prospective Bidders.

As prescribed in 14.201-6(c)(2), insert the following provision:

14.201-6 Solicitation provisions.

(c) Insert the following provisions in invitations for bids:

(2) 52.214-6, Explanation to Prospective Bidders.

Explanation to Prospective Bidders (Apr 1984)

Any prospective bidder desiring an explanation or interpretation of the solicitation, drawings, specifications, etc., must request it in writing soon enough to allow a reply to reach all prospective bidders before the submission of their bids. Oral explanations or instructions given before the award of a contract will not be binding. Any information given a prospective bidder concerning a solicitation will be furnished promptly to all other prospective bidders as an amendment to the solicitation, if that information is necessary in submitting bids or if the lack of it would be prejudicial to other prospective bidders.

(End of provision)

52.214-7 Late Submissions, Modifications, and Withdrawals of Bids.

As prescribed in 14.201-6(c)(3), insert the following provision:

14.201-6 Solicitation provisions.

(c) Insert the following provisions in invitations for bids:

(3) 52.214-7, Late Submissions, Modifications, and Withdrawals of Bids.

Late Submissions, Modifications, and Withdrawals of Bids (Nov 1999)

(a) Bidders are responsible for submitting bids, and any modifications or withdrawals, so as to reach the Government office designated in the invitation for bids (IFB) by the time specified in the IFB. If no time is specified in the IFB, the time for receipt is 4:30 p.m., local time, for the designated Government office on the date that bids are due.

(b)(1) Any bid, modification, or withdrawal received at the Government office designated in the IFB after the exact time specified for receipt of bids is "late" and will not be considered unless it is received before award is made, the Contracting Officer determines that accepting the late bid would not unduly delay the acquisition; and--

(i) If it was transmitted through an electronic commerce method authorized by the IFB, it was received at the initial point of entry to the Government infrastructure not later than 5:00 p.m. one working day prior to the date specified for receipt of bids; or

(ii) There is acceptable evidence to establish that it was received at the Government installation designated for receipt of bids and was under the Government's control prior to the time set for receipt of bids.

(2) However, a late modification of an otherwise successful bid that makes its terms more favorable to the Government, will be considered at any time it is received and may be accepted.

(c) Acceptable evidence to establish the time of receipt at the Government installation includes the time/date stamp of that installation on the bid wrapper, other documentary evidence of receipt maintained by the installation, or oral testimony or statements of Government personnel.

(d) If an emergency or unanticipated event interrupts normal Government processes so that bids cannot be received at the Government office designated for receipt of bids by the exact time specified in the IFB and urgent Government requirements preclude amendment of the IFB, the time specified for receipt of bids will be deemed to be extended to the same time of day specified in the solicitation on the first work day on which normal Government processes resume.

(e) Bids may be withdrawn by written notice received at any time before the exact time set for receipt of bids. If the IFB authorizes facsimile bids, bids may be withdrawn via facsimile received at any time before the exact time set for receipt of bids, subject to the conditions specified in the provision at 52.214-31, Facsimile Bids. A bid may be withdrawn in person by a bidder or its authorized representative if, before the exact time set for receipt of bids, the identity of the person requesting withdrawal is established and the person signs a receipt for the bid.

(End of provision)

52.214-8 [Reserved]

52.214-9 Failure to Submit Bid.

As prescribed in 14.201-6(e)(1), insert the following provision in invitations for bids:

14.201-6 Solicitation provisions.

(e) Insert in invitations for bids, except those for construction, the provisions at--

(1) 52.214-9, Failure to Submit Bid, except when using electronic data interchange methods not requiring solicitation mailing lists; and

Failure to Submit Bid (Jul 1995)

Recipients of this solicitation not responding with a bid should not return this solicitation, unless it specifies otherwise. Instead, they should advise the issuing office by letter, postcard, or established electronic commerce methods, whether they want to receive future solicitations for similar requirements. If a recipient does not submit a bid and does not notify the issuing office that future solicitations are desired, the recipient's name may be removed from the applicable mailing list.

(End of provision)

52.214-10 Contract Award--Sealed Bidding.

As prescribed in 14.201-6(e)(2), insert the following provision:

14.201-6 Solicitation provisions.

(e) Insert in invitations for bids, except those for construction, the provisions at:

(2) 52.214-10, Contract Award--Sealed Bidding.

Contract Award--Sealed Bidding (Jul 1990)

(a) The Government will evaluate bids in response to this solicitation without discussions and will award a contract to the responsible bidder whose bid, conforming to the solicitation, will be most advantageous to the Government considering only price and the price-related factors specified elsewhere in the solicitation.

(b) The Government may--

(1) Reject any or all bids;

(2) Accept other than the lowest bid; and

(3) Waive informalities or minor irregularities in bids received.

(c) The Government may accept any item or group of items of a bid, unless the bidder qualifies the bid by specific limitations. Unless otherwise provided in the Schedule, bids may be submitted for quantities less than those specified. The Government reserves the right to make an award on any item for a quantity less than the quantity offered, at the unit prices offered, unless the bidder specifies otherwise in the bid.

(d) A written award or acceptance of a bid mailed or otherwise furnished to the successful bidder within the time for acceptance specified in the bid shall result in a binding contract without further action by either party.

(e) The Government may reject a bid as nonresponsive if the prices bid are materially unbalanced between line items or subline items. A bid is materially unbalanced when it is based on prices significantly less than cost for some work and prices which are significantly overstated in relation to cost for other work, and if there is a reasonable doubt that the bid will result in the lowest overall cost to the Government even though it may be the low evaluated bid, or if it is so unbalanced as to be tantamount to allowing an advance payment.

(End of provision)

52.214-11 [Reserved]

52.214-12 Preparation of Bids.

As prescribed in 14.201-6(f), insert the following provision:

14.201-6 Solicitation provisions.

(f) Insert in invitations for bids to which the uniform contract format applies, the provision at 52.214-12, Preparation of Bids.

Preparation of Bids (Apr 1984)

(a) Bidders are expected to examine the drawings, specifications, Schedule, and all instructions. Failure to do so will be at the bidder's risk.

(b) Each bidder shall furnish the information required by the solicitation. The bidder shall sign the bid and print or type its name on the Schedule and each continuation sheet on which it makes an entry. Erasures or other changes must be initialed by the person signing the bid. Bids signed by an agent shall be accompanied by evidence of that agent's authority, unless that evidence has been previously furnished to the issuing office.

(c) For each item offered, bidders shall (1) show the unit price, including, unless otherwise specified, packaging, packing, and preservation and (2) enter the extended price for the quantity of each item offered in the "Amount" column of the Schedule. In case of discrepancy between a unit price and an extended price, the unit price will be presumed to be correct, subject, however, to correction to the same extent and in the same manner as any other mistake.

(d) Bids for supplies or services other than those specified will not be considered unless authorized by the solicitation.

(e) Bidders must state a definite time for delivery of supplies or for performance of services, unless otherwise specified in the solicitation.

(f) Time, if stated as a number of days, will include Saturdays, Sundays, and holidays.

(End of provision)

52.214-13 Telegraphic Bids.

As prescribed in 14.201-6(g)(1), insert the following provision:

14.201-6 Solicitation provisions.

(g)(1) Insert the provision at 52.214-13, Telegraphic Bids, in invitations for bids if the contracting officer decides to authorize telegraphic bids.

(2) Use the provision with its Alternate I in invitations for bids that are for perishable subsistence, and when the contracting officer considers that offerors will be unwilling to provide acceptance periods long enough to allow written confirmation.

Telegraphic Bids (Apr 1984)

(a) Bidders may submit telegraphic bids as responses to this solicitation. These responses must arrive at the place, and by the time, specified in the solicitation.

(b) Telegraphic bids shall refer to this solicitation and include the items or subitems, quantities, unit prices, time and place of delivery, all representations and other information required by this solicitation, and a statement of agreement with all the terms, conditions, and provisions of the invitation for bids.

(c) Telegraphic bids that fail to furnish required representations or information, or that reject any of the terms, conditions, and provisions of the solicitation, may be excluded from consideration.

(d) Bidders must promptly sign and submit complete copies of the bids in confirmation of their telegraphic bids.

(e) The term "telegraphic bids," as used in this provision, includes mailgrams.

(End of provision)

Alternate I (Nov 1988). As prescribed in 14.201-6(g)(2), substitute the following for paragraph (d) of the basic clause:

(d) Written confirmation of telegraphic bids is not required.

(End of Alternate provision)

52.214-14 Place of Performance--Sealed Bidding.

As prescribed in 14.201-6(h), insert the following provision:

14.201-6 Solicitation provisions.

(h) Insert the provision at 52.214-14, Place of Performance--Sealed Bidding, in invitations for bids except those in which the place of performance is specified by the Government.

Place of Performance--Sealed Bidding (Apr 1985)

(a) The bidder, in the performance of any contract resulting from this solicitation, * intends, * does not intend [*check applicable box*] to use one or more plants or facilities located at a different address from the address of the bidder as indicated in this bid.

(b) If the bidder checks "intends" in paragraph (a) above, it shall insert in the spaces provided below the required information:

(End of provision)

52.214-15 Period for Acceptance of Bids.

As prescribed in 14.201-6(i), insert the following provision:

14.201-6 Solicitation provisions.

(i) Insert the provision at 52.214-15, Period for Acceptance of Bids, in invitations for bids (IFB's) that are not issued on SF 33 or SF 1447 except IFB's--

(1) For construction work; or

(2) In which the Government specifies a minimum acceptance period.

Period for Acceptance of Bids (Apr 1984)

In compliance with the solicitation, the bidder agrees, if this bid is accepted within _______ calendar days (60 calendar days unless a different period is inserted by the bidder) from the date specified in the solicitation for receipt of bids, to furnish any or all items upon which prices are bid at the price set opposite each item, delivered at the designated point(s), within the time specified in the Schedule.

(End of provision)

52.214-16 Minimum Bid Acceptance Period.

As prescribed in 14.201-6(j), insert the following provision in invitations for bids, except for construction, if the contracting officer determines that a minimum acceptance period must be specified:

14.201-6 Solicitation provisions.

(j) Insert the provision at 52.214-16, Minimum Bid Acceptance Period, in invitations for bids, except for construction, if the contracting officer determines that a minimum acceptance period must be specified.

Minimum Bid Acceptance Period (Apr 1984)

(a) "Acceptance period," as used in this provision, means the number of calendar days available to the Government for awarding a contract from the date specified in this solicitation for receipt of bids.

(b) This provision supersedes any language pertaining to the acceptance period that may appear elsewhere in this solicitation.

(c) The Government requires a minimum acceptance period of __________ calendar days [*the Contracting Officer shall insert the number of days*].

(d) In the space provided immediately below, bidders may specify a longer acceptance period than the Government's minimum requirement.

The bidder allows the following acceptance period: ______________ calendar days.

(e) A bid allowing less than the Government's minimum acceptance period will be rejected.

(f) The bidder agrees to execute all that it has undertaken to do, in compliance with its bid, if that bid is accepted in writing within--

(1) The acceptance period stated in paragraph (c) of this clause; or

(2) Any longer acceptance period stated in paragraph (d) of this clause.

(End of provision)

52.214-17 [Reserved]

52.214-18 Preparation of Bids--Construction.

As prescribed in 14.201-6(l), insert the following provision:

14.201-6 Solicitation provisions.

(l) Insert the provision at 52.214-18, Preparation of Bids--Construction, in invitations for bids for construction work.

Preparation of Bids--Construction (Apr 1984)

(a) Bids must be--

(1) Submitted on the forms furnished by the Government or on copies of those forms, and

(2) *Manually signed.* The person signing a bid must initial each erasure or change appearing on any bid form.

(b) The bid form may require bidders to submit bid prices for one or more items on various bases, including--

(1) Lump sum bidding;

(2) Alternate prices;

(3) Units of construction; or

(4) Any combination of subparagraphs (1) through (3) above.

(c) If the solicitation requires bidding on all items, failure to do so will disqualify the bid. If bidding on all items is not required, bidders should insert the words "no bid" in the space provided for any item on which no price is submitted.

(d) Alternate bids will not be considered unless this solicitation authorizes their submission.

(End of provision)

52.214-19 Contract Award--Sealed Bidding-- Construction.

As prescribed in 14.201-6(m), insert the following provision:

14.201-6 Solicitation provisions.

(m) Insert the provision at 52.214-19, Contract Award--Sealed Bidding-- Construction, in all invitations for bids for construction work.

Contract Award--Sealed Bidding--Construction (Aug 1996)

(a) The Government will evaluate bids in response to this solicitation without discussions and will award a contract to the responsible bidder whose bid, conforming to the solicitation, will be most advantageous to the Government, considering only price and the price-related factors specified elsewhere in the solicitation.

(b) The Government may reject any or all bids, and waive informalities or minor irregularities in bids received.

(c) The Government may accept any item or combination of items, unless doing so is precluded by a restrictive limitation in the solicitation or the bid.

(d) The Government may reject a bid as nonresponsive if the prices bid are materially unbalanced between line items or subline items. A bid is materially unbalanced when it is based on prices significantly less than cost for some work and prices which are significantly overstated in relation to cost for other work, and if there is a reasonable doubt that the bid will result in the lowest overall cost to the Government even though it may be the low evaluated bid, or if it is so unbalanced as to be tantamount to allowing an advance payment.

(End of provision)

52.214-20 Bid Samples.

As prescribed in 14.201-6(o)(1), insert the following provision in invitations for bids if bid samples are required:

14.201-6 Solicitation provisions.

(o)(1) Insert the provision at 52.214-20, Bid Samples, in invitations for bids if bid samples are required.

(2) If it appears that the conditions in 14.202-4(f)(1) will apply and the contracting officer anticipates granting waivers thereunder and--

(i) If the nature of the required product does not necessitate limiting the grant of a waiver to a product produced at the same plant in which the product previously acquired or tested was produced, use the provision with its Alternate I; or

(ii) If the nature of the required product necessitates limiting the grant of a waiver to a product produced at the same plant in which the product previously acquired or tested was produced, use the provision with its Alternate II.

(3) See 14.202-4(f)(2) regarding waiving the requirement for all bidders.

14.202-4 Bid samples.

(f) *Waiver of requirement for bid samples*.

(1) The requirement for furnishing bid samples may be waived when a bidder offers a product previously or currently being contracted for or tested by the Government and found to comply with specification requirements conforming in every material respect with those in the current invitation for bids. When the requirement may be waived, see 14.201-6(o)(2).

(2) Where samples required by a Federal, Military, or other formal specification are not considered necessary and a waiver of the sample requirements of the specification has been authorized, a statement shall be included in the invitation that notwithstanding the requirements of the specification, samples will not be required.

Bid Samples (Apr 1984)

(a) "Bid samples" are item sample submissions required of bidders to show those characteristics of the offered products that cannot adequately be described by specifications or purchase descriptions (*e.g.*, balance, facility of use, or pattern).

(b) Bid samples, required elsewhere in this solicitation, must be furnished as part of the bid and must be received by the time specified for receipt of bids. Failure to furnish samples on time will require rejection of the bid, except that a late sample sent by mail may be considered under the Late Submissions, Modifications, and Withdrawals of Bids provision of this solicitation.

(c) Bid samples will be tested or evaluated to determine compliance with all the characteristics listed for examination in this solicitation. Failure of these samples to conform to the required characteristics will require rejection of the bid. Products delivered under any resulting contract must conform to--

(1) The approved sample for the characteristics listed for test or evaluation; and

(2) The specifications for all other characteristics.

(d) Unless otherwise specified in the solicitation, bid samples shall be--

(1) Submitted at no expense to the Government; and

(2) Returned at the bidder's request and expense, unless they are destroyed during preaward testing.

(End of provision)

***Alternate I (Apr 1984*).** If it appears that the conditions in 14.202-4(f)(1) will apply and the Contracting Officer anticipates granting waivers thereunder, and if the nature of the required product *does not* necessitate limiting the grant of a waiver to a product produced at the same plant in which the product previously acquired or tested was produced, add the following paragraph (e) to the basic provision:

(e) At the discretion of the Contracting Officer, the requirement for furnishing bid samples may be waived for a bidder if--

(1) The bid states that the offered product is the same as a product offered by the bidder to the _______________ [*as appropriate, the Contracting Officer shall designate the contracting office or an alternate activity or office*]; and

(2) The Contracting Officer determines that the previously offered product was accepted or tested and found to comply with specification and other requirements for technical acceptability conforming in every material respect with those in this solicitation.

(End of Alternate provision)

***Alternate II (APR 1984*).** If it appears that the conditions in 14.202-4(f)(1) will apply and the contracting officer anticipates granting waivers thereunder, and if the nature of the required product necessitates limiting the grant of a waiver to a product produced at the same plant in which the product previously acquired or tested was produced, add the following paragraph (e) to the basic provision:

(e) At the discretion of the Contracting Officer, the requirements for furnishing bid samples may be waived for a bidder if--

(1) The bid states that the offered product is the same as a product offered by the bidder to the _________ [*as appropriate, the Contracting Officer shall designate the contracting office or an alternate activity or office*] on a previous acquisition;

(2) The Contracting Officer determines that the previously offered product was accepted or tested and found to comply with specification and other requirements for technical acceptability conforming in every material respect with those of this solicitation; and

(3) The product offered under this solicitation will be produced under a resulting contract at the same plant in which the previously acquired or tested product was produced.

(End of Alternate provision)

52.214-21 Descriptive Literature.

As prescribed in 14.201-6(p)(1), insert the following provision:

14.201-6 Solicitation provisions.

(p)(1) Insert the provision at 52.214-21, Descriptive Literature, in invitations for bids if--

(i) Descriptive literature is required to evaluate the technical acceptability of an offered product; and

(ii) The required information will not be readily available unless it is submitted by bidders.

(2) Use the basic clause with its Alternate I if the possibility exists that the contracting officer may waive the requirement for furnishing descriptive literature for a bidder offering a previously supplied product that meets specification requirements of the current solicitation.

(3) See 14.202-5(e)(2) regarding waiving the requirement for all bidders.

Descriptive Literature (Apr 1984)

(a) "Descriptive literature" means information (*e.g.*, cuts, illustrations, drawings, and brochures) that is submitted as part of a bid. Descriptive literature is required to establish, for the purpose of evaluation and award, details of the product offered that are specified elsewhere in the solicitation and pertain to significant elements such as (1) design; (2) materials; (3) components; (4) performance characteristics; and (5) methods of manufacture, assembly, construction, or operation. The term includes only information required to determine the technical acceptability of the offered product. It does not include other information such as that used in determining the responsibility of a prospective Contractor or for operating or maintaining equipment.

(b) Descriptive literature, required elsewhere in this solicitation, must be (1) identified to show the item(s) of the offer to which it applies and (2) received by the time specified in this solicitation for receipt of bids. Failure to submit descriptive literature on time will require rejection of the bid, except that late descriptive literature sent by mail may be considered under the Late Submissions, Modifications, and Withdrawals of Bids provision of this solicitation.

(c) The failure of descriptive literature to show that the product offered conforms to the requirements of this solicitation will require rejection of the bid.

(End of provision)

Alternate I (May 1999). As prescribed in 14.201-6(p)(2), add the following paragraphs (d) and (e) to the basic provision.

(d) The Contracting Officer may waive the requirement for furnishing descriptive literature if the bidder has supplied a product the same as that required by this solicitation under a prior contract. A bidder that requests a waiver of this requirement shall provide the following information:

Prior contract number _________________________

Date of prior contract _________________________

Contract line item number of product supplied ____

Name and address of government activity to which delivery was made ____________________________

Date of final delivery of product supplied _________

(e) Bidders must submit bids on the basis of required descriptive literature or on the basis of a previously supplied product under paragraph (d) above. A bidder submitting a bid on one of these two bases may not elect to have its bid considered on the alternative basis after the time specified for receipt of bids. A bidder's request for a waiver under paragraph (d) above will be disregarded if that bidder has submitted the descriptive literature required under this solicitation.

(End of Alternate provision)

52.214-22 Evaluation of Bids for Multiple Awards.

As prescribed in 14.201-6(q), insert the following provision:

14.201-6 Solicitation provisions.

(q) Insert the provision at 52.214-22, Evaluation of Bids for Multiple Awards, in invitations for bids if the contracting officer determines that multiple awards might be made if doing so is economically advantageous to the Government.

Evaluation of Bids for Multiple Awards (Mar 1990)

In addition to other factors, bids will be evaluated on the basis of advantages and disadvantages to the Government that might result from making more than one award (multiple awards). It is assumed, for the purpose of evaluating bids, that $500 would be the administrative cost to the Government for issuing and administering each contract awarded under this solicitation, and individual awards will be for the items or combinations of items that result in the lowest aggregate cost to the Government, including the assumed administrative costs.

(End of provision)

52.214-23 Late Submissions, Modifications, Revisions, and Withdrawals of Technical Proposals under Two-Step Sealed Bidding.

As prescribed in 14.201-6(r), insert the following provision:

14.201-6 Solicitation provisions.

(r) Insert the provision at 52.214-23, Late Submissions, Modifications, Revisions, and Withdrawals of Technical Proposals under Two-Step Sealed Bidding, in solicitations for technical proposals in step one of two-step sealed bidding.

Late Submissions, Modifications, Revisions, and Withdrawals of Technical Proposals under Two-Step Sealed Bidding (Nov 1999)

(a) Bidders are responsible for submitting technical proposals, and any modifications or revisions, so as to reach the Government office designated in the request for technical proposals by the time specified in the invitation for bids (IFB). If no time is specified in the IFB, the time for receipt is 4:30 p.m., local time, for the designated Government office on the date that bids or revisions are due.

(b)(1) Any technical proposal under step one of two-step sealed bidding, modification, revision, or withdrawal of such proposal received at the Government office designated in the request for technical proposals after the exact time specified for receipt will not be considered unless the Contracting Officer determines that accepting the late technical proposal would not unduly delay the acquisition; and--

(i) If it was transmitted through an electronic commerce method authorized by the request for technical proposals, it was received at the initial point of entry to the Government infrastructure not later than 5:00 p.m. one working day prior to the date specified for receipt of proposals; or

(ii) There is acceptable evidence to establish that it was received at the Government installation designated for receipt of offers and was under the Government's control prior to the time set for receipt; or

(iii) It is the only proposal received and it is negotiated under Part 15 of the Federal Acquisition Regulation.

(2) However, a late modification of an otherwise successful proposal that makes its terms more favorable to the Government will be considered at any time it is received and may be accepted.

(c) Acceptable evidence to establish the time of receipt at the Government installation includes the time/date stamp of that installation on the technical proposal wrapper, other documentary evidence of receipt maintained by the installation, or oral testimony or statements of Government personnel.

(d) If an emergency or unanticipated event interrupts normal Government processes so that technical proposals cannot be received at the Government office designated for receipt of technical proposals by the exact time specified in the request for technical proposals, and urgent Government requirements preclude amendment of the request for technical proposals, the time specified for receipt of technical proposals will be deemed to be extended to the same time of day specified in the request for technical proposals on the first work day on which normal Government processes resume.

(e) Technical proposals may be withdrawn by written notice received at any time before the exact time set for receipt of technical proposals. If the request for technical proposals authorizes facsimile technical proposals, they may be withdrawn via facsimile received at any time before the exact time set for receipt of proposals, subject to the conditions specified in the provision at 52.214-31, Facsimile Bids. A technical proposal may be withdrawn in person by a bidder or its authorized representative if, before the exact time set for receipt of technical proposals, the identity of the person requesting withdrawal is established and the person signs a receipt for the technical proposal.

(End of provision)

52.214-24 Multiple Technical Proposals.

As prescribed in 14.201-6(s), insert the following provision:

14.201-6 Solicitation provisions.

(s) Insert the provision at 52.214-24, Multiple Technical Proposals, in solicitations for technical proposals in step one of two-step sealed bidding if the contracting officer permits the submission of multiple technical proposals.

Multiple Technical Proposals (Apr 1984)

In the first step of this two-step acquisition, solicited sources are encouraged to submit multiple technical proposals presenting different basic approaches. Each technical proposal submitted will be separately evaluated and the submitter will be notified as to its acceptability.

(End of provision)

52.214-25 Step Two of Two-Step Sealed Bidding.

As prescribed in 14.201-6(t), insert the following provision:

14.201-6 Solicitation provisions.

(t) Insert the provision at 52.214-25, Step Two of Two-Step Sealed Bidding, in invitations for bids issued under step two of two-step sealed bidding.

Step Two of Two-Step Sealed Bidding (Apr 1985)

(a) This invitation for bids is issued to initiate step two of two-step sealed bidding under Subpart 14.5 of the Federal Acquisition Regulation.

(b) The only bids that the Contracting Officer may consider for award of a contract are those received from bidders that have submitted acceptable technical proposals in step one of this acquisition under _________________[*the Contracting Officer shall insert the identification of the step-one request for technical proposals*].

(c) Any bidder that has submitted multiple technical proposals in step one of this acquisition may submit a separate bid on each technical proposal that was determined to be acceptable to the Government.

(End of provision)

52.214-26 Audit and Records--Sealed Bidding.

As prescribed in 14.201-7(a), insert the following clause:

14.201-7 Contract clauses.

(a) When contracting by sealed bidding, the contracting officer shall insert the clause at 52.214-26, Audit and Records--Sealed Bidding, in solicitations and contracts if the contract amount is expected to exceed the threshold at 15.403-4(a)(1) for submission of cost or pricing data.

Audit and Records--Sealed Bidding (Oct 1997)

(a) As used in this clause, "records" includes books, documents, accounting procedures and practices, and other data, regardless of type and regardless of whether such items are in written form, in the form of computer data, or in any other form.

(b) *Cost or pricing data*. If the Contractor has been required to submit cost or pricing data in connection with the pricing of any modification to this contract, the Contracting Officer, or an authorized representative of the Contracting Officer, in order to evaluate the accuracy, completeness, and currency of the cost or pricing data, shall have the right to examine and audit all of the Contractor's records, including computations and projections, related to--

(1) The proposal for the modification;

(2) The discussions conducted on the proposal(s), including those related to negotiating;

(3) Pricing of the modification; or

(4) Performance of the modification.

(c) *Comptroller General*. In the case of pricing any modification, the Comptroller General of the United States, or an authorized representative, shall have the same rights as specified in paragraph (b) of this clause.

(d) *Availability*. The Contractor shall make available at its office at all reasonable times the materials described in paragraph (b) of this clause, for examination, audit, or reproduction, until 3 years after final payment under this contract, or for any other period specified in Subpart 4.7 of the Federal Acquisition Regulation (FAR). FAR Subpart 4.7, Contractor Records Retention, in effect on the date of this contract, is incorporated by reference in its entirety and made a part of this contract.

(1) If this contract is completely or partially terminated, the records relating to the work terminated shall be made available for 3 years after any resulting final termination settlement.

(2) Records pertaining to appeals under the Disputes clause or to litigation or the settlement of claims arising under or relating to the performance of this contract shall be made available until disposition of such appeals, litigation, or claims.

(e) The Contractor shall insert a clause containing all the provisions of this clause, including this paragraph (e), in all subcontracts expected to exceed the threshold in FAR 15.403-4(a)(1) for submission of cost or pricing data.

(End of clause)

52.214-27 Price Reduction for Defective Cost or Pricing Data--Modifications--Sealed Bidding.

As prescribed in 14.201-7(b), insert the following clause:

14.201-7 Contract clauses.

(b)(1) When contracting by sealed bidding, the contracting officer shall insert the clause at 52.214-27, Price Reduction for Defective Cost or Pricing Data--Modifications--Sealed Bidding, in solicitations and contracts if the contract amount is expected to exceed the threshold for submission of cost or pricing data at 15.403-4(a)(1).

(2) In exceptional cases, the head of the contracting activity may waive the requirement for inclusion of the clause in a contract with a foreign government or agency of that government. The authorizations for the waiver and the reasons for granting it shall be in writing.

15.403-4 Requiring cost or pricing data (10 U.S.C. 2306a and 41 U.S.C. 254b).

(a)(1) Cost or pricing data shall be obtained only if the contracting officer concludes that none of the exceptions in 15.403-1(b) applies. However, if the contracting officer has sufficient information available to determine price reasonableness, then a waiver under the exception at 15.403-1(b)(4) should be considered. The threshold for obtaining cost or pricing data is $500,000.

Unless an exception applies, cost or pricing data are required before accomplishing any of the following actions expected to exceed the current threshold or, in the case of existing contracts, the threshold specified in the contract:

(i) The award of any negotiated contract (except for undefinitized actions such as letter contracts).

(ii) The award of a subcontract at any tier, if the contractor and each higher-tier subcontractor have been required to furnish cost or pricing data (but see waivers at 15.403-1(c)(4)).

(iii) The modification of any sealed bid or negotiated contract (whether or not cost or pricing data were initially required) or any subcontract covered by paragraph (a)(1)(ii) of this subsection. Price adjustment amounts shall consider both increases and decreases (*e.g.*, a $150,000 modification resulting from a reduction of $350,000 and an increase of $200,000 is a pricing adjustment exceeding $500,000). This requirement does not apply when unrelated and separately priced changes for which cost or pricing data would not otherwise be required are included for administrative convenience in the same modification. Negotiated final pricing actions (such as termination settlements and total final price agreements for fixed-price incentive and redeterminable contracts) are contract modifications requiring cost or pricing data if the total final price agreement for such settlements or agreements exceeds the pertinent threshold set forth at paragraph (a)(1) of this subsection, or the partial termination settlement plus the estimate to complete the continued portion of the contract exceeds the pertinent threshold set forth at paragraph (a)(1) of this subsection (see 49.105(c)(15)).

(2) Unless prohibited because an exception at 15.403-1(b) applies, the head of the contracting activity, without power of delegation, may authorize the contracting officer to obtain cost or pricing data for pricing actions below the pertinent threshold in paragraph (a)(1) of this subsection, provided the action exceeds the simplified acquisition threshold. The head of the contracting activity shall justify the requirement for cost or pricing data. The documentation shall include a written finding that cost or pricing data are necessary to determine whether the price is fair and reasonable and the facts supporting that finding.

(b) When cost or pricing data are required, the contracting officer shall require the contractor or prospective contractor to submit to the contracting officer (and to have any subcontractor or prospective subcontractor submit to the prime contractor or appropriate subcontractor tier) the following in support of any proposal:

(1) The cost or pricing data.

(2) A certificate of current cost or pricing data, in the format specified in 15.406-2, certifying that to the best of its knowledge and belief, the cost or pricing data were accurate, complete, and current as of the date of agreement on price or, if applicable, an earlier date agreed upon between the parties that is as close as practicable to the date of agreement on price.

(c) If cost or pricing data are requested and submitted by an offeror, but an exception is later found to apply, the data shall not be considered cost or pricing data as defined in 15.401 and shall not be certified in accordance with 15.406-2.

(d) The requirements of this subsection also apply to contracts entered into by an agency on behalf of a foreign government.

Price Reduction for Defective Cost or Pricing Data--Modifications--Sealed Bidding (Oct 1997)

(a) This clause shall become operative only for any modification to this contract involving aggregate increases and/or decreases in costs, plus applicable profits, expected to exceed the threshold for the submission of cost or pricing data at FAR 15.403-4(a)(1), except that this clause does not apply to a modification if an exception under FAR 15.403-1(b) applies.

(b) If any price, including profit, negotiated in connection with any modification under this clause, was increased by any significant amount because (1) the Contractor or a subcontractor furnished cost or pricing data that were not complete, accurate, and current as certified in its Certificate of Current Cost or Pricing Data, (2) a subcontractor or prospective subcontractor furnished the Contractor cost or pricing data that were not complete, accurate, and current as certified in the Contractor's Certificate of Current Cost or Pricing Data, or (3) any of these parties furnished data of any description that were not accurate, the price shall be reduced accordingly and the contract shall be modified to reflect the reduction. This right to a price reduction is limited to that resulting from defects in data relating to modifications for which this clause becomes operative under paragraph (a) of this clause

(c) Any reduction in the contract price under paragraph (b) of this clause due to defective data from a prospective subcontractor that was not subsequently awarded the subcontract shall be limited to the amount, plus applicable overhead and profit markup, by which (1) the actual subcontract or (2) the actual cost to the Contractor, if there was no subcontract, was less than the prospective subcontract cost estimate submitted by the Contractor; provided, that the actual subcontract price was not itself affected by defective cost or pricing data.

(d)(1) If the Contracting Officer determines under paragraph (b) of this clause that a price or cost reduction should be made, the Contractor agrees not to raise the following matters as a defense:

(i) The Contractor or subcontractor was a sole source supplier or otherwise was in a superior bargaining position and thus the price of the contract would not have been modified even if accurate, complete, and current cost or pricing data had been submitted.

(ii) The Contracting Officer should have known that the cost or pricing data in issue were defective even though the Contractor or subcontractor took no affirmative action to bring the character of the data to the attention of the Contracting Officer.

(iii) The contract was based on an agreement about the total cost of the contract and there was no agreement about the cost of each item procured under the contract.

(iv) The Contractor or subcontractor did not submit a Certificate of Current Cost or Pricing Data.

(2)(i) Except as prohibited by subdivision (d)(2)(ii) of this clause, an offset in an amount determined appropriate by the Contracting Officer based upon the facts shall be allowed against the amount of a contract price reduction if--

(A) The Contractor certifies to the Contracting Officer that, to the best of the Contractor's knowledge and belief, the Contractor is entitled to the offset in the amount requested; and

(B) The Contractor proves that the cost or pricing data were available before the date of agreement on the price of the contract (or price of the modification) and that the data were not submitted before such date.

(ii) An offset shall not be allowed if--

(A) The understated data was known by the Contractor to be understated when the Certificate of Current Cost or Pricing Data was signed; or

(B) The Government proves that the facts demonstrate that the contract price would not have increased in the amount to be offset even if the available data had been submitted before the date of agreement on price.

(e) If any reduction in the contract price under this clause reduces the price of items for which payment was made prior to the date of the modification reflecting the price reduction, the Contractor shall be liable to and shall pay the United States at the time such overpayment is repaid--

(1) Simple interest on the amount of such overpayment to be computed from the date(s) of overpayment to the Contractor to the date the Government is repaid by the Contractor at the applicable underpayment rate effective for each quarter prescribed by the Secretary of the Treasury under 26 U.S.C. 6621(a)(2); and

(2) A penalty equal to the amount of the overpayment, if the Contractor or subcontractor knowingly submitted cost or pricing data which were incomplete, inaccurate, or noncurrent.

(End of clause)

52.214-28 Subcontractor Cost or Pricing Data-- Modifications-- Sealed Bidding.

As prescribed in 14.201-7(c), insert the following clause in solicitations and contracts:

14.201-7 Contract clauses.

(c)(1) When contracting by sealed bidding, the contracting officer shall insert the clause at 52.214-28, Subcontractor Cost or Pricing Data--Modifications-- Sealed Bidding, in solicitations and contracts if the contract amount is expected to exceed the threshold for submission of cost or pricing data at 15.403-4(a)(1).

(2) In exceptional cases, the head of the contracting activity may waive the requirement for inclusion of the clause in a contract with a foreign government or agency of that government. The authorizations for the waiver and the reasons for granting it shall be in writing.

15.403-4 Requiring cost or pricing data (10 U.S.C. 2306a and 41 U.S.C. 254b).

(a)(1) Cost or pricing data shall be obtained only if the contracting officer concludes that none of the exceptions in 15.403-1(b) applies. However, if the contracting officer has sufficient information available to determine price reasonableness, then a waiver under the exception at 15.403-1(b)(4) should be considered. The threshold for obtaining cost or pricing data is $500,000. Unless an exception applies, cost or pricing data are required before accomplishing any of the following actions expected to exceed the current threshold or, in the case of existing contracts, the threshold specified in the contract:

(i) The award of any negotiated contract (except for undefinitized actions such as letter contracts).

(ii) The award of a subcontract at any tier, if the contractor and each higher-tier subcontractor have been required to furnish cost or pricing data (but see waivers at 15.403-1(c)(4)).

(iii) The modification of any sealed bid or negotiated contract (whether or not cost or pricing data were initially required) or any subcontract covered by paragraph (a)(1)(ii) of this subsection. Price adjustment amounts shall consider both increases and decreases (*e.g.*, a $150,000 modification resulting from a reduction of $350,000 and an increase of $200,000 is a pricing adjustment exceeding $500,000). This requirement does not apply when unrelated and separately priced changes for which cost or pricing data would not otherwise be required are included for administrative convenience in the same modification. Negotiated final pricing actions (such as termination settlements and total final price agreements for fixed-price incentive and redeterminable contracts) are contract modifications requiring cost or pricing data if the total final price agreement for such settlements or agreements exceeds the pertinent threshold set forth at paragraph (a)(1) of this subsection, or the partial termination settlement plus the estimate to complete the continued portion of the contract exceeds the pertinent threshold set forth at paragraph (a)(1) of this subsection (see 49.105(c)(15)).

(2) Unless prohibited because an exception at 15.403-1(b) applies, the head of the contracting activity, without power of delegation, may authorize the contracting officer to obtain cost or pricing data for pricing actions below the pertinent threshold in paragraph (a)(1) of this subsection, provided the action exceeds the simplified acquisition threshold. The head of the contracting activity shall justify the requirement for cost or pricing data. The documentation shall include a written finding that cost or pricing data are necessary to determine whether the price is fair and reasonable and the facts supporting that finding.

(b) When cost or pricing data are required, the contracting officer shall require the contractor or prospective contractor to submit to the contracting officer (and to have any subcontractor or prospective subcontractor submit to the prime contractor or appropriate subcontractor tier) the following in support of any proposal:

(1) The cost or pricing data.

(2) A certificate of current cost or pricing data, in the format specified in 15.406-2, certifying that to the best of its knowledge and belief, the cost or pricing data were accurate, complete, and current as of the date of agreement on price or, if applicable, an earlier date agreed upon between the parties that is as close as practicable to the date of agreement on price.

(c) If cost or pricing data are requested and submitted by an offeror, but an exception is later found to apply, the data shall not be considered cost or pricing data as defined in 15.401 and shall not be certified in accordance with 15.406-2.

(d) The requirements of this subsection also apply to contracts entered into by an agency on behalf of a foreign government.

Subcontractor Cost or Pricing Data--Modifications--Sealed Bidding (Oct 1997)

(a) The requirements of paragraphs (b) and (c) of this clause shall--

(1) Become operative only for any modification to this contract involving aggregate increases and/or decreases in costs, plus applicable profits, expected to exceed the threshold for submission of cost or pricing data at FAR 15.403-4(a)(1); and

(2) Be limited to such modifications.

(b) Before awarding any subcontract expected to exceed the threshold for submission of cost or pricing data at FAR 15.403-4(a)(1), on the date of agreement on price or the date of award, whichever is later; or before pricing any subcontract modifications involving aggregate increases and/or decreases in costs, plus applicable profits, expected to exceed the threshold for submission of cost or pricing data at FAR 15.403-4(a)(1), the Contractor shall require the subcontractor to submit cost or pricing data (actually or by specific identification in writing), unless an exception under FAR 15.403-1(b) applies.

(c) The Contractor shall require the subcontractor to certify in substantially the form prescribed in FAR subsection 15.4062 that, to the best of its knowledge and belief, the data submitted under paragraph (b) of this clause were accurate, complete, and current as of the date of agreement on the negotiated price of the subcontract or subcontract modification.

(d) The Contractor shall insert the substance of this clause, including this paragraph (d), in each subcontract that, when entered into, exceeds the threshold for submission of cost or pricing data at FAR 15.403-4(a)(1).

(End of clause)

52.214-29 Order of Precedence--Sealed Bidding.

As prescribed in 14.201-7(d), insert the following clause:

14.201-7 Contract clauses.

(d) When contracting by sealed bidding the contracting officer shall insert the clause at 52.214-29, Order of Precedence--Sealed Bidding, in solicitations and contracts to which the uniform contract format applies.

Order of Precedence--Sealed Bidding (Jan 1986)

Any inconsistency in this solicitation or contract shall be resolved by giving precedence in the following order:

(a) The Schedule (excluding the specifications);

(b) Representations and other instructions;

(c) Contract clauses;

(d) Other documents, exhibits, and attachments; and

(e) The specifications.

(End of clause)

52.214-30 Annual Representations and Certifications--Sealed Bidding.

As prescribed in 14.201-6(u), insert the following provision:

14.201-6 Solicitation provisions.

(u) Insert the provision at 52.214-30, Annual Representations and Certifications--Sealed Bidding, in invitations for bids if annual representations and certifications are used (see 14.213).

14.213 Annual submission of representations and certifications.

(a) Submission of offeror representations and certifications on an annual basis, as an alternative to submission in each solicitation, may be authorized by agencies subject to the requirements of this section. The decision to use annual representations and certifications shall be made in accordance with agency procedures.

(b) In accordance with agency procedures, each contracting office utilizing annual representations and certifications shall establish procedures and assign responsibilities for centrally requesting, receiving, storing, verifying and updating offeror's annual submissions. Generally, the representations and certifications shall be effective for a period of 1 year from date of signature.

(c) The contracting officer shall not include in individual solicitations the full text of provisions that are contained in the annual representations and certifications.

(d) Offerors shall make changes that affect only one solicitation by completing the appropriate section of the provision at 52.214-30, Annual Representations and Certifications--Sealed Bidding.

Annual Representations and Certifications--Sealed Bidding (Jan 1997)

The bidder has (check the appropriate block):

* (a) Submitted to the contracting office issuing this solicitation, annual representations and certifications dated ______________________________ [*insert date of signature on submission*], which are incorporated herein by reference, and are current, accurate, and complete as of the date of this bid, except as follows [*insert changes that affect only this solicitation; if "none," so state*]: _____________

* (b) Enclosed its annual representations and certifications.

(End of provision)

52.214-31 Facsimile Bids.

As prescribed in 14.201-6(v), insert the following provision:

14.201-6 Solicitation provisions.

(v) Insert the provision at 52.214-31, Facsimile Bids, in solicitations if facsimile bids are authorized (see 14.202-7).

14.202-7 Facsimile bids.

(a) Unless prohibited or otherwise restricted by agency procedures, contracting officers may authorize facsimile bids (see 14.201-

6(v)). In determining whether or not to authorize facsimile bids, the contracting officer shall consider factors such as--

(1) Anticipated bid size and volume;

(2) Urgency of the requirement;

(3) Frequency of price changes;

(4) Availability, reliability, speed, and capacity of the receiving facsimile equipment; and

(5) Adequacy of administrative procedures and controls for receiving, identifying, recording, and safeguarding facsimile bids, and ensuring their timely delivery to the bids opening location.

(b) If facsimile bids are authorized, contracting officers may, after the date set for bid opening, request the apparently successful offeror to provide the complete, original signed bid.

Facsimile Bids (Dec 1989)

(a) *Definition*. "Facsimile bid," as used in this solicitation, means a bid, modification of a bid, or withdrawal of a bid that is transmitted to and received by the Government via electronic equipment that communicates and reproduces both printed and handwritten material.

(b) Bidders may submit facsimile bids as responses to this solicitation. These responses must arrive at the place and by the time, specified in the solicitation.

(c) Facsimile bids that fail to furnish required representations or information or that reject any of the terms, conditions, and provisions of the solicitation may be excluded from consideration.

(d) Facsimile bids must contain the required signatures.

(e) The Government reserves the right to make award solely on the facsimile bid. However, if requested to do so by the Contracting Officer, the apparently successful bidder agrees to promptly submit the complete original signed bid.

(f) Facsimile receiving data and compatibility characteristics are as follows:

(1) Telephone number of receiving facsimile equipment:

____

(2) Compatibility characteristics of receiving facsimile equipment (*e.g*., make and model number, receiving speed, communications protocol):

____

(g) If the bidder chooses to transmit a facsimile bid, the Government will not be responsible for any failure attributable to the transmission or receipt of the facsimile bid including, but not limited to, the following:

(1) Receipt of garbled or incomplete bid.

(2) Availability or condition of the receiving facsimile equipment.

(3) Incompatibility between the sending and receiving equipment.

(4) Delay in transmission or receipt of bid.

(5) Failure of the bidder to properly identify the bid.

(6) Illegibility of bid.

(7) Security of bid data.

(End of provision)

52.214-32--52.214-33 [Reserved]

52.214-34 Submission of Offers in the English Language.

As prescribed in 14.201-6(w), insert the following provision:

14.201-6 Solicitation provisions.

(w) Insert the provision at 52.214-34, Submission of Offers in the English Language, in solicitations that include any of the clauses prescribed in 25.1101 or 25.1102. It may be included in other solicitations when the contracting officer decides that it is necessary.

25.1101 Acquisition of supplies.

The following provisions and clauses apply to the acquisition of supplies and the acquisition of services involving the furnishing of supplies.

(a)(1) Insert the clause at 52.225-1, Buy American Act--Balance of Payments Program--Supplies, in solicitations and contracts with a value exceeding $2,500 but not exceeding $25,000; and in solicitations and contracts with a value exceeding $25,000, if none of the clauses prescribed in paragraphs (b) and (c) of this section apply, except if--

(i) The solicitation is restricted to domestic end products in accordance with Subpart 6.3;

(ii) The acquisition is for supplies for use within the United States and an exception to the Buy American Act applies (*e.g.*, nonavailability or public interest); or

(iii) The acquisition is for supplies for use outside the United States and an exception to the Balance of Payments Program applies.

(2) Insert the provision at 52.225-2, Buy American Act--Balance of Payments Program Certificate, in solicitations containing the clause at 52.225-1.

(b)(1)(i) Insert the clause at 52.225-3, Buy American Act--North American Free Trade Agreement--Israeli Trade Act--Balance of Payments Program, in solicitations and contracts with a value exceeding $25,000 but less than $177,000, unless--

(A) The acquisition is for the acquisition of supplies, or for services involving the furnishing of supplies, for use outside the United States, and the value of the acquisition is less than the simplified acquisition threshold; or

(B) The acquisition is exempt from the North American Free Trade Agreement and the Israeli Trade Act (see 25.401). For acquisitions of agencies not subject to the Israeli Trade Act (see 25.406), see agency regulations.

(ii) If the acquisition value exceeds $25,000 but is less than $50,000, use the clause with its Alternate I.

(iii) If the acquisition value is $50,000 or more but less than $54,372, use the clause with its Alternate II.

(2)(i) Insert the provision at 52.225-4, Buy American Act--North American Free Trade Agreement--Israeli Trade Act--Balance of Payments Program Certificate, in solicitations containing the clause at 52.225-3.

(ii) If the acquisition value exceeds $25,000 but is less than $50,000, use the provision with its Alternate I.

(iii) If the acquisition value is $50,000 or more but less than $54,372, use the provision with its Alternate II.

(c)(1) Insert the clause at 52.225-5, Trade Agreements, in solicitations and contracts valued at $177,000 or more, if the Trade Agreements Act applies (see 25.401 and 25.403) and the agency has determined that the restrictions of the Buy American Act or Balance of Payments Program are not applicable to U.S.-made end products, unless the acquisition is to be awarded and performed outside the United States in support of a contingency operation or a humanitarian or peacekeeping operation and does not exceed the increased simplified acquisition threshold of $200,000. If the agency has not made such a determination, the contracting officer must follow agency procedures.

(2) Insert the provision at 52.225-6, Trade Agreements Certificate, in solicitations containing the clause at 52.225-5.

(d) Insert the provision at 52.225-7, Waiver of Buy American Act for Civil Aircraft and Related Articles, in solicitations for civil aircraft and related articles (see 25.407), if the acquisition value is less than $177,000.

(e) Insert the clause at 52.225-8, Duty-Free Entry, in solicitations and contracts for supplies that may be imported into the United States and for which duty-free entry may be obtained in accordance with 25.903(a), if the value of the acquisition--

(1) Exceeds $100,000; or

(2) Is $100,000 or less, but the savings from waiving the duty is anticipated to be more than the administrative cost of waiving the duty. When used for acquisitions valued at $100,000 or less, the contracting officer may modify paragraphs (b)(1) and (i)(2) of the clause to reduce the dollar figure.

25.1102 Acquisition of construction.

(a) Insert the clause at 52.225-9, Buy American Act--Balance of Payments Program--Construction Materials, in solicitations and contracts for construction valued at less than $6,806,000.

(1) List in paragraph (b)(2) of the clause all foreign construction material excepted from the requirements of the Buy American Act.

(2) If the head of the agency determines that a higher percentage is appropriate, substitute the higher evaluation percentage in paragraph (b)(3)(i) of the clause.

(b)(1) Insert the provision at 52.225-10, Notice of Buy American Act/Balance of Payments Program Requirement--Construction Materials, in solicitations containing the clause at 52.225-9.

(2) If insufficient time is available to process a determination regarding the inapplicability of the Buy American Act or Balance of Payments Program prior to receipt of offers, use the provision with its Alternate I.

(c) Insert the clause at 52.225-11, Buy American Act--Balance of Payments Program--Construction Materials under Trade Agreements, in solicitations and contracts valued at $6,806,000 or more.

(1) List in paragraph (b)(3) of the clause all foreign construction material excepted from the requirements of the Buy American Act, other than designated country or NAFTA country construction material.

(2) If the head of the agency determines that a higher percentage is appropriate, substitute the higher evaluation percentage in paragraph (b)(4)(i) of the clause.

(3) For acquisitions valued at $6,806,000 or more, but less than $7,068,419, use the clause with its Alternate I.

(d)(1) Insert the provision at 52.225-12, Notice of Buy American Act/Balance of Payments Program Requirement--Construction Materials under Trade Agreements, in solicitations containing the clause at 52.225-11.

(2) If insufficient time is available to process a determination regarding the inapplicability of the Buy American Act or Balance of Payments Program before receipt of offers, use the provision with its Alternate I.

(3) For acquisitions valued at $6,806,000 or more, but less than $7,068,419, use the clause with its Alternate II.

Submission of Offers in the English Language (Apr 1991)

Offers submitted in response to this solicitation shall be in the English language. Offers received in other than English shall be rejected.

(End of provision)

52.214-35 Submission of Offers in U.S. Currency.

As prescribed in 14.201-6(x), insert the following provision:

14.201-6 Solicitation provisions.

(x) Insert the provision at 52.214-35, Submission of Offers in U.S. Currency, in solicitations that include any of the clauses prescribed in 25.1101 or 25.1102, unless the contracting officer includes the clause at 52.225-17, Evaluation of Foreign Currency Offers, as prescribed in 25.1103(d). It may be included in other solicitations when the contracting officer decides that it is necessary.

25.1101 Acquisition of supplies.

The following provisions and clauses apply to the acquisition of supplies and the acquisition of services involving the furnishing of supplies.

(a)(1) Insert the clause at 52.225-1, Buy American Act--Balance of Payments Program--Supplies, in solicitations and contracts with a value exceeding $2,500 but not exceeding $25,000; and in solicitations and contracts with a value exceeding $25,000, if none of the clauses prescribed in paragraphs (b) and (c) of this section apply, except if--

(i) The solicitation is restricted to domestic end products in accordance with Subpart 6.3;

(ii) The acquisition is for supplies for use within the United States and an exception to the Buy American Act applies (*e.g.*, nonavailability or public interest); or

(iii) The acquisition is for supplies for use outside the United States and an exception to the Balance of Payments Program applies.

(2) Insert the provision at 52.225-2, Buy American Act--Balance of Payments Program Certificate, in solicitations containing the clause at 52.225-1.

(b)(1)(i) Insert the clause at 52.225-3, Buy American Act--North American Free Trade Agreement--Israeli Trade Act--Balance of Payments Program, in solicitations and contracts with a value exceeding $25,000 but less than $177,000, unless--

(A) The acquisition is for the acquisition of supplies, or for services involving the furnishing of supplies, for use outside the United States, and the value of the acquisition is less than the simplified acquisition threshold; or

(B) The acquisition is exempt from the North American Free Trade Agreement and the Israeli Trade Act (see 25.401). For acquisitions of agencies not subject to the Israeli Trade Act (see 25.406), see agency regulations.

(ii) If the acquisition value exceeds $25,000 but is less than $50,000, use the clause with its Alternate I.

(iii) If the acquisition value is $50,000 or more but less than $54,372, use the clause with its Alternate II.

(2)(i) Insert the provision at 52.225-4, Buy American Act--North American Free Trade Agreement--Israeli Trade Act--Balance of Payments Program Certificate, in solicitations containing the clause at 52.225-3.

(ii) If the acquisition value exceeds $25,000 but is less than $50,000, use the provision with its Alternate I.

(iii) If the acquisition value is $50,000 or more but less than $54,372, use the provision with its Alternate II.

(c)(1) Insert the clause at 52.225-5, Trade Agreements, in solicitations and contracts valued at $177,000 or more, if the Trade Agreements Act applies (see 25.401 and 25.403) and the agency has determined that the restrictions of the Buy American Act or Balance of Payments Program are not applicable to U.S.-made end products, unless the acquisition is to be awarded and performed outside the United States in support of a contingency operation or a humanitarian or peacekeeping operation and does not exceed the increased simplified acquisition threshold of $200,000. If the agency has not made such a determination, the contracting officer must follow agency procedures.

(2) Insert the provision at 52.225-6, Trade Agreements Certificate, in solicitations containing the clause at 52.225-5.

(d) Insert the provision at 52.225-7, Waiver of Buy American Act for Civil Aircraft and Related Articles, in solicitations for civil aircraft and related articles (see 25.407), if the acquisition value is less than $177,000.

(e) Insert the clause at 52.225-8, Duty-Free Entry, in solicitations and contracts for supplies that may be imported into the United States and for which duty-free entry may be obtained in accordance with 25.903(a), if the value of the acquisition--

(1) Exceeds $100,000; or

(2) Is $100,000 or less, but the savings from waiving the duty is anticipated to be more than the administrative cost of waiving the duty. When used for acquisitions valued at $100,000 or less, the contracting officer may modify paragraphs (b)(1) and (i)(2) of the clause to reduce the dollar figure.

25.1102 Acquisition of construction.

(a) Insert the clause at 52.225-9, Buy American Act--Balance of Payments Program--Construction Materials, in solicitations and contracts for construction valued at less than $6,806,000.

(1) List in paragraph (b)(2) of the clause all foreign construction material excepted from the requirements of the Buy American Act.

(2) If the head of the agency determines that a higher percentage is appropriate, substitute the higher evaluation percentage in paragraph (b)(3)(i) of the clause.

(b)(1) Insert the provision at 52.225-10, Notice of Buy American Act/Balance of Payments Program Requirement--Construction Materials, in solicitations containing the clause at 52.225-9.

(2) If insufficient time is available to process a determination regarding the inapplicability of the Buy American Act or Balance of Payments Program prior to receipt of offers, use the provision with its Alternate I.

(c) Insert the clause at 52.225-11, Buy American Act--Balance of Payments Program--Construction Materials under Trade Agreements, in solicitations and contracts valued at $6,806,000 or more.

(1) List in paragraph (b)(3) of the clause all foreign construction material excepted from the requirements of the Buy American Act, other than designated country or NAFTA country construction material.

(2) If the head of the agency determines that a higher percentage is appropriate, substitute the higher evaluation percentage in paragraph (b)(4)(i) of the clause.

(3) For acquisitions valued at $6,806,000 or more, but less than $7,068,419, use the clause with its Alternate I.

(d)(1) Insert the provision at 52.225-12, Notice of Buy American Act/Balance of Payments Program Requirement--Construction Materials under Trade Agreements, in solicitations containing the clause at 52.225-11.

(2) If insufficient time is available to process a determination regarding the inapplicability of the Buy American Act or Balance of Payments Program before receipt of offers, use the provision with its Alternate I.

(3) For acquisitions valued at $6,806,000 or more, but less than $7,068,419, use the clause with its Alternate II.

25.1103 Other provisions and clauses.

(a) *Restrictions on certain foreign purchases*. Insert the clause at 52.225-13, Restrictions on Certain Foreign Purchases, in solicitations and contracts with a value exceeding $2,500.

(b) *Translations*. Insert the clause at 52.225-14, Inconsistency Between English Version and Translation of Contract, in solicitations and contracts if anticipating translation into another language.

(c) *Sanctions*. (1) Except as provided in paragraph (c)(2) of this section, insert the clause at--

(i) 52.225-15, Sanctioned European Union Country End Products, in solicitations and contracts for supplies valued at less than $177,000; or

(ii) 52.225-16, Sanctioned European Union Country Services, in solicitations and contracts for services--

(A) Listed in 25.601(a)(3)(i); or

(B) Valued at less than $177,000.

(2) Do not insert the clauses in paragraph (c)(1) of this section in--

(i) Solicitations issued and contracts awarded by a contracting activity located outside of the United States, provided the supplies will be used or the services will be performed outside of the United States;

(ii) Purchases at or below the simplified acquisition threshold awarded using simplified acquisition procedures;

(iii) Total small business set-asides;

(iv) Contracts in support of U.S. national security interests;

(v) Contracts for essential spare, repair, or replacement parts available only from sanctioned EU member states; or

(vi) Contracts for which the head of the agency has made a determination in accordance with 25.602(b).

(d) *Foreign currency offers*. Insert the provision at 52.225-17, Evaluation of Foreign Currency Offers, in solicitations that permit the use of other than a specified currency. Insert in the provision the source of the rate to be used in the evaluation of offers.

Submission of Offers in U.S. Currency (Apr 1991)

Offers submitted in response to this solicitation shall be in terms of U.S. dollars. Offers received in other than U.S. dollars shall be rejected.

(End of provision)

BIBLIOGRAPHY

48 Code of Federal Regulations. Federal Acquisition Regulations. Office of Management and Budget. US Printing Office Washington.

Ackerman, Mark S. and Thomas W. Malone. "Answer Garden: A Tool for Growing Organizational Memory." *Association of Computing Machinery* (1990): 31-39.

Allen, Maryellen Mott. "The myth of Intelligent Agents." *ONLINE* (November/December 2000): 45-51.

Anderson, James D. "Standards for Indexing: Revising the American National Standard Guidelines Z39.4." *Journal of American Society of Information Science* 45(8) (1994): 628-636.

Armacost, Robert L. and Jamshid C. Hosseini. "Identification of Determinant Attributes Using the Analytic Hierarchy Process." *Journal of the Academy of Marketing Science* 22(4) (1994): 383-392.

Barry, Carol L. "User-Defined Relevance Criteria: An Exploratory Study." *Journal of American Society of Information Science* 45(3) (1994): 149-159.

Bates, Marcia J. "The Design of Browsing and Berry-picking Techniques for the Online Search Interface." *Online Review* 13(5) (1989): 407-424.

Belkin, Nicholas J. "Anomalous States of Knowledge as a Basis for Information Retrieval." *Canadian Journal of Information Science* 5 (1980): 133-144.

Belkin, Nicholas J., R. N. Oddy, and H. M. Brooks. "ASK For Information Retrieval: Part I. Background and Theory." *Journal of Documentation* 38-2 (June 1982): 61-71.

________. "ASK For Information Retrieval: Part II. Results of a Design Study." *Journal of Documentation* 38-3 (September 1982): 145-163.

Belkin, Nicholas J., and Alina Vickery. *Interaction in Information Systems: A Review of Research from Document Retrieval to Knowledge-Based Systems*. Boston Spa: British Library, 1985.

Belkin, Nicholas J.and W. Bruce Croft. "Information Filtering and Information Retrieval: Two Sides of the Same Coin?" *Communications of the ACM* 35(12) (1992): 29-38.

Bench-Capon, T. J. M. *Knowledge Representation: An Approach to Artificial Intelligence*. San Diego, CA: Academic Press, Inc., 1990.

Bollmann-Sdorra, Peter and Vjay V. Raghavan. "On the Delusiveness of Adopting a Common Space for Modeling IR Objects: Are Queries Documents?" *Journal of American Society of Information Science 44(10)* (1993): 579-587.

Bordogna, Gloria and Gabriella Pasi. "A Fuzzy Linguistic Approach Generalizing Boolean Information Retrieval: A Model and Its Evaluation." *Journal of American Society of Information Science* 44(2) (1993): 70-82.

Bowen, T. F., G. Gopal, G. Herman, T. Hickey, K. C. Lee, W. H. Mansfield, J. Raitz, and A. Weinrib. "The Datacycle Architecture." *Communications of the ACM* 35(12) (1992): 71-81.

Brooks, Terrence A. "All the Right Descriptors: A Test of the Strategy of Unlimited Aliasing." *Journal of American Society of Information Science* 44(3) (1993): 137-147.

Bruce, Bertram. "Case Systems for Natural Language." *Artificial Intelligence* 6 (1975): 327-360.

Bruce, Harry W. "A Cognitive View of the Situational Dynamism of User-Centered Relevance Estimation." *Journal of American Society of Information Science* 45(3) (1994): 142-148.

Bunge, Charles A. "Planning, Goals, and Objectives for the Reference Department." *Reference Quarterly* (Spring 1984): 306-315.

Cleverdon, Cyril. "The Cranfield Tests on Index Language Devices." in *ASLIB Proceedings* 19(6), 1967.

Cohen, Paul R. and Rick Kjeldsen. "Information Retrieval by Constrained Spreading Activation in Semantic Networks." *Information Processing and Management* 23(4) (1987): 255-268.

Cole, Charles. "Operationalizing the Notion of Information as a Subjective Construct." *Journal of American Society of Information Science* 45(7) (1994): 465-476.

Conklin, Jeff. "Hypertext: An Introduction and Survey." *Computer* 9(20) (1987): 17-41.

Cook, Kenneth H. "A Threshold Model of Relevance Decisions." *Information Processing and Management* 11(5) (1975): 125-135.

Cooper, W. S. "A Definition of Relevance for Information Retrieval." *Information Storage and Retrieval* 7(1) (1971): 19-37.

Cuadra, Carlos A. and Robert V. Katter. "Opening the Black Box of Relevance." *Journal of Documentation* 23(4) (1967): 291-303.

Deerwester, S., S. T. Dumais, G. W. Furnas, T. R. Landauer, and R. Harshman. "Indexing by Latent Semantic Analysis." *Journal of American Society of Information Science* 41(6) (1990): 391-407.

Derr, Richard L. "Information Seeking Expressions of Users." *Journal of American Society of Information Science* 35 (1984): 124-28.

________. "Questions: Definition, Structure, and Classification." *Reference Quarterly*, *23* (1984): 1-11.

Dervin, Brenda. "Information as a User Construct: The Relevance of Perceived Information Needs to a Synthesis and Interpretation." in *Knowledge Structure and Use: Implications for Synthesis and Interpretation*. Philadelphia: Temple University Press, 1983. 155-183.

________. "Communication Gaps and Inequities: Moving Toward a Reconceptualization." in Dervin, Brenda; Voigt, Melvin, eds. *Progress in Communication Sciences.* Norwood: Ablex, 1980.

Dervin, Brenda and P. Dewdney. “Neutral Questioning: A New Approach to the Reference Interview.” *Reference Quarterly* 25 (Summer 1986): 506-513.

Dervin, Brenda, Thomas L. Jacobson, and Michael S. Nilan. “Measuring Aspects of Information Seeking: A Test of a Quantitative / Qualitative Methodology.” *Information Management* (1982): 419-444

Ding, Wei and Gary Marchionini. *A comparative study of Web Search Service Performance*. College Park, MD: University of Maryland/CLIS, 1995.

Doty, D. Harold and William H. Glick. “Typologies as a Unique Form of Theory Building: Toward Improved Understanding and Modeling.” *Academy of Management Review* 19(2) (1992): 230-251.

Duces, Brigitte. “The Role of Information in the Adoption of Agricultural Innovations” Ph.D. diss., University of Maryland/CLIS, 1985.

Ellis, D. “Modeling the Information-seeking Patterns of Academic Researchers: A Grounded Theory Approach.” *The Library Quarterly* 63(4) (1993): 469-486.

Etzioni, Oren and Daniel A. Weld. “Softbot-based Interface to the Internet.” *Communications of the ACM* 37(7) (1994): 72-76.

Fidel, Raya. “Online Searching Styles: A Case-Study-Based Model of Searching Behavior.” *Journal of American Society of Information Science* 35(4) (1984): 211-221.

________. “Searchers' Selection of Search Keys: II. Controlled Vocabulary or Free-Text Searching.” *Journal of American Society of Information Science* 42(7) (1991): 501-14.

________. “Searchers' Selection of Search Keys: III. Searching Styles.” *Journal of American Society of Information Science* 42(7) (1991): 515-527.

________. “User-Centered Indexing.” *Journal of American Society of Information Science 45(8)* (1994): 572-576.

Fidel, Raya and Dagobert Soergel. "Factors Affecting Online Bibliographic Retrieval: A Conceptual Framework for Research." *Journal of American Society of Information Science* 34(3) (1983): 163-180.

Fikes, Richard and Tom Kehler. "The Role of Frame-based Representation in Reasoning." *Communications of the ACM* 28(9) (1985): 904-920.

Filmore, Charles J. "Topics in Lexical Semantics," in *Current Issues in Linguistics Theory*. Bloomington, IN: Indiana University Press, 1977. 76-138.

Foltz, Peter W. and Susan T. Dumais. "Personalized Information Delivery: An Analysis of Information Filtering Methods." *Communications of the ACM* 35(12) (1992): 51-60.

Frakes, William B. and Ricardo Baeza-Yates, eds. *Information Retrieval: Data Structures and Algorithms*. Englewood Cliffs, NJ: Prentice Hall, 1992.

Froehlich, Thomas J. "Relevance Reconsidered-Towards an Agenda for the 21st Century: Introduction to Special Topic Issue on Relevance Research." *Journal of American Society of Information Science* 45(3) (1994): 124-134.

Gauch, Susan and John Smith. "An Expert System for Automatic Query Reformation." *Journal of American Society of Information Science* 44(3) (1993): 124-136.

Genescreth, Michael R. and Steven P. Ketchpel. "Software Agents." *Communications of the ACM* 37(7) (1994): 48-53, 147.

Giarratano, Joseph, and Gary Riley. *Expert Systems: Principles and Programming*. Boston: PWS-KENT Publishing Company, 1989.

Gisolfi, Antonio. "Classifying through an Algebraic Fuzzy Structure: The Relevance of the Attributes." *International Journal of Intelligent Systems* 20 (1995): 715-734.

Glazier, Jack D. and Ronald R. Powell, eds. *Qualitative Research in Information Management.* Englewood, CO: Libraries Unlimited, 1992.

Goker, Ayse Safiye. “An Investigation into the Application of Machine Learning in Information Retrieval.” Ph.D. diss., London, City University, Department of Information Science, 1994.

Goldberg, David, David Nichols, Brian M. Oki, and Douglas Terry. “Using Collaborative Filtering to Weave an Information Tapestry.” *Communications of the ACM* 35(12) (1992): 61-70.

Green, Rebecca. “The Expression of Syntagmatic Relationships in Frame-Based Indexing.” Ph.D. diss., University of Maryland/CLIS, 1989.

________. *Topical Relevance Relationships. I. Why Topic Matching Fails.* University of Maryland/CLIS, 1994.

Griffin, Stephen & Khurshid Ahmad. “Archiving Knowledge Before and After the Interview.” *Knowledge Acquisition Using Conceptual Graph Theory Proceedings, ICCS 94 Workshop*, University of Maryland, College Park, Maryland, 1994.

Guba, E. G. “Criteria for Assessing the Trustworthiness of Naturalistic Inquires.” *Educational Communication and Technology Journal* 29(2) (1981): 75-91.

Guba, E. G. and Y. S. Lincoln. “Epistemological and Methodological Bases of Naturalistic Inquiry.” *Educational Communication and Technology Journal* 30(4) (1982): 233-252.

Hayes-Roth, Frederick. “Rule-Based Systems.” *Communications of the ACM* 28(9) (1985): 921-932.

Hersh, William. “Relevance and Retrieval Evaluation: Perspectives from Medicine.” *Journal of American Society of Information Science* 45(3) (1994): 201-206.

Howard, Dara Lee. “Pertinence as Reflected in Personal Constructs.” *Journal of American Society of Information Science* 45(3) (1994): 172-185.

Hunter, Geoffrey. *Metalogic: An Introduction to the Metatheory of Standard First Order Logic.* Berkeley, CA: University of California Press, 1971.

Jacobson, Thomas L. “Sense-Making in a Database Environment.” *Information Processing & Management* 27 (1991): 647-657.

Janes, Joseph W. “Other People's Judgments: A Comparison of Others' Judgments of Document Relevance, Topicality, and Utility.” *Journal of American Society of Information Science* 45(3) (1994): 160-171.

Javelin, Kalervo and Timo Niemi. “Deductive Information Retrieval Based on Classifications.” *Journal of American Society of Information Science* 44(10) (1993): 557-578.

Kautz, Henry A., Bart Selman, and Michael Coen. “Bottom-up Design of Software Agents.” *Communications of the ACM* 37(7) (1994): 143-146.

Kemp, D. A. “Relevance, Pertinence, and Information System Development.” *Information Storage and Retrieval* 10(2) (1974): 37-47.

King, Donald W. and Edward C. Bryant. *The Evaluation of Information Services and Products.* Washington, DC: Information Resources Press, 1971.

Klein, Katherine J., Fred Dansereau, and Rosalie J. Hall. “Levels Issues in Theory Development, Data Collection, and Analysis.” *Academy of Management Review* 19(2) (1994): 195-229.

Kuhlthau, Carol C. “Inside the Search Process: Information Seeking from the User's Perspective.” *Journal of American Society of Information Science* 42 (1991): 361-371.

Kyng, Morten. “Making Representations Work.” *Communications of the ACM* 38(9) (1995): 46-55.

Lancaster, F. Wilfrid. “MEDLARS: Report on the Evaluation of its Operating Efficiency.” *American Documentation* 20(2) (1969): 8-19.

Lindsay, Peter H. and Donald H. Norman. *Human Information Processing. Introduction to Psychology.* New York: Academic Press, 1972.

Loeb, Shoshana. “Architecting Personalized Delivery of Multimedia Information.” *Communications of the ACM* 35(12) (1992): 39-50.

Mackay, Wendy E., Thomas W. Malone, Kevin Crowston, Ramana Rao, David Rosenblatt, and Stuart Card. "How Do Experienced Users Information Lens Users Use Rules?" *Proceedings of the ACM Conference on Human Factors in Computing Systems,* Austin, Texas, April 30-May 4, 1989.

Maes, Patti. "Agents That Reduce Work and Information Overload." *Communications of the ACM* 37(7) (1994): 31-40, 146.

Malone, Thomas W., Kenneth R. Grant, and Franklyn A. Turbak. "The Information Lens: An Intelligent System for Information Sharing in Organizations." *Proceedings of the ACM conference CHI '86*, 1986, 1-8.

Marchionini, Gary. *Information Seeking in Electronic Environments*. New York: Cambridge University Press, 1995.

Marchionini, Gary and H. Maurer. "The Roles of DLs in Teaching and Learning." *Communications of the ACM* 38(4) (1995): 67-75.

Marchionini, Gary, Diane Barlow, and Linda Hill. "Extending Retrieval Strategies to Networked Environments: Old Ways, New Ways, and a Critical Look at WAIS." *Journal of American Society of Information Science* 45(8) (1994): 561-564.

Mayer, Richard E. "Rule Induction: Thinking as Hypothesis Testing." in *Thinking, Problem Solving, Cognition, 2d ed.* New York: W. H. Freman, 1992.

McCarn, Davis B. and Charles R. Stein. "Intelligence Systems Evaluation." in *Electronic Handling of Information.* Washington, DC/London: Thompson Academic Press, 1967.

McCleary, Hunter. "Filtered Information Services: A Revolutionary New Product or a New Marketing Strategy?" *ONLINE* 18(4) (1994): 33-42.

Michalski, Ryszard S. and Robert E. Stepp. "Learning from Observation: Conceptual clustering." in *Machine Learning: An Artificial Intelligence Approach.* Ryszard S. Michalski, Jame G. Carbonell and Tom M. Mitchell, eds. Palo Alto, CA: Tioga Publishing Company, 1983.

Michel, Dee Andy. “What Is Used During Cognitive Processing in Information Retrieval and Library Searching? Eleven sources of Search Information.” *Journal of American Society of Information Science* 45(7) (1994): 498-514.

Milstead, Jessica L. “Needs for Research in Indexing.” *Journal of American Society of Information Science* 45(8) (1994): 577-582.

National Institute on Alcohol Abuse and Alcoholism and Center for Substance Abuse Prevention. *The Alcohol and Other Drugs Thesaurus: A Guide to Concepts and Terminology in Substance Abuse and Addiction*, 2d ed., Rockville, MD: National Institute on Alcohol Abuse and Alcoholism and Center for Substance Abuse Prevention, 1993.

Nielson, Paul E. “SOAR/IFOR: Intelligent Agents for Air Simulation and Control.” *Proceedings of the 1995 Winter Simulation Conference.* C. Alexopolos, K. Kang, W. R. Lilegdon, and D. Goldsman, eds. (1995): 620-625.

Oard, Douglas W. “The State of the Art in Text Filtering.” *User Modeling and User Adapted Interaction* 7(3) (1997): 141-178.

Ottaviani, J. S. “The Fractal Nature of Relevance: A Hypothesis.” *Journal of American Society of Information Science* 45(4) (1994): 263-272.

Paijmana, Hans. “Comparing the Document Representations of Two IR Systems: CLARIT and TOPIC.” *Journal of American Society of Information Science* 44(8) (1993): 383-392.

Pao, Miranda Lee. “Relevance Odds of Retrieval Overlaps From Seven Search Fields.” *Information Processing & Management* 30(3) (1994): 305-314.

Papazoglou, M. P. “Unraveling the Semantics of Conceptual Schemas.” *Communications of the ACM* 38(9) (1995): 80-94.

Park, Taemin Kim. “Toward a Theory of User-Based Relevance: A Call for a New Paradigm of Inquiry.” *Journal of American Society of Information Science* 45(3) (1994): 135-141.

Pollitt, Steven. “CANSEARCH: An Expert Systems Approach to Document Retrieval.” *Information Processing & Management* 23(2) (1987): 119-138.

Purtill, Richard L. *Logic for Philosophers*. New York: Harper & Row, 1971.

Ram, Suda. “Deriving Functional Dependencies from the Entity-Relationship Model.” *Communications of the ACM* 38(9) (1995): 95-107.

Rao, R., J. O. Pedersen, M. A. Hearst, J. D. MacKinlay, S. K. Card, L. Masinter, P. K. Halvorsen, and G. G. Robertson. “Rich interaction in the DL.” *Communications of the ACM* 38(4) (1995): 29-39.

Rau, Lisa. “Knowledge Organization and Access in a Conceptual Information System.” *Information Processing and Management* 23(4) (1987): 269-283.

Resnick, Paul, Neophytos Iacovou, Mitesh Suchak, Peter Bergstrom, and John Riedl. “GroupLens: An Open Architecture for Collaborative Filtering of Netnews.” *Association of Computing Machinery* (1994): 175-186.

Riecken, Doug. “A Conversation with Marvin Minsky about Agents.” *Communications of the ACM* 37(7) (1994): 23-29.

Rosenfeld, Louis B. and Maurita P. Holland. “Automated Filtering of Internet Postings.” *ONLINE* (May 1994): 27-30.

Saaty, Thomas L. *The Analytic Hierarchy Process*. New York: McGraw-Hill, 1980.

Salton, G. and M. McGill. *Introduction to Modern Information Retrieval*. New York: McGraw-Hill, 1983.

Saracevic, Tefko, Paul Kantor, Alice Y. Chamis, and Donna Trivison. “A Study of Information Seeking and Retrieving: I. Background and Methodology; II. Users, Questions, and Effectiveness; and III. Searchers, Searches, and Overlap.” *Journal of American Society of Information Science* 39 (1988): 161-216.

Saracevic, Tefko, ed. *Introduction to Information Science*. New York: R. R. Bowker Co., 1970.

Schamber, Linda. “Relevance and Information Behavior.” *Annual Review of Information Science and Technology* 29(1) (1994): 1-48.

Schamber, Linda, Michael B. Eisenberg, and Michael S. Nilan. “A Re-examination of Relevance: Toward a Dynamic, Situational Definition.” *Information Processing and Management* 26(6) (1990): 755-776.

Schank, R. C. and R. P. Abelson, “Scripts, Plans and Knowledge.” in *Thinking: Readings in Cognitive Science.* Cambridge, England: Cambridge University Press, 1977.

Shneiderman, Ben. “Dynamic Queries for Visual Information Seeking.” *Technical Report of University of Maryland*, 1994, *CS-TR-3022*.

Shuldberg, H. Kelly, Melissa Macpherson, Pete Humphrey, and Jamil Corley. “Distilling Information from Text: The EDS *TemplateFiller* System.” *Journal of American Society of Information Science* 44(9) (1993): 493-507.

Skemp, Richard R. *The Psychology of Learning Mathematics*. Hammondsworth, UK: Penguin Books, Inc., 1986.

Smithson, S. “Information Retrieval Evaluation in Practice: A Case Study Approach.” *Information Processing and Management* 30(2) (1994): 205-221.

Soergel, Dagobert. *Indexing Languages and Thesauri.* Los Angeles: Melville Publishing Company, 1974.

________. *Organizing Information: Principles of Data Base and Retrieval Systems*. Orlando, FL: Academic Press, 1985.

________. “Improving Access to Food and Nutrition Data: A Language for the Description of Foods in Databases.” *Language* (1990): 1-16.

________. "Indexing and Retrieval Performance: The Logical Evidence." *Journal of American Society of Information Science* 45(8) (1994): 589-599.

________. *Information Structure Management: A Unified Framework for Indexing and Searching Database, Expert, Information-Retrieval, and Hypermedia Systems.* Medford, NJ: Learned Information, 1994.

________. "Software Support for Thesaurus Construction and Display." in *Proceedings of the 5th ASIS SIG/CR Classification Research Workshop, October 16, 1994 in conjunction with 57th ASIS Annual Meeting.* Alexandria, VA: American Society for Information Science SIG/CR: 157-184.

________. "SemWeb: Proposal for an Open, Multifunctional, Multilingual, System for Integrated Access to Knowledge Base About Concepts and Terminology." *Proceedings of the Fourth International ISDO Conference*, 15-18 July 1996, Washington, DC. 165-173.

Sowa, John F. *Knowledge Representation: Logical, Philosophical, and Computational Foundations, preliminary ed. ICCS 94*. College Park: University of Maryland, 1994.

Strauss, A. and J. Corbin. "Coding Procedures: Open Coding." in *Basics of Qualitative Research.* Thousand Oaks, CA: Sage, 1990.

Su, Louise T. "The Relevance of Recall and Precision in User Evaluation." *Journal of American Society of Information Science* 45(3) (1994): 207-217.

Sutton, Stuart A. "The Role of Attorney Mental Models of Law in Case Relevance Determinations: An Exploratory Analysis." *Journal of American Society of Information Science* 45(3) (1994): 186-200.

Swigger, Keith. "Questions in Library and Information Science." *Library and Information Science Research* 7 (1985): 369-83.

Taylor. R. S. "The Process of Asking Questions." *American Documentation* 13(4) (1962): 391-396.

Tenopir, Carol and Pamela Cahn. "Target and Freestyle: DIALOG and Mead Join the Relevance Ranks." *ONLINE* (May 1994): 31-47.

Vaill, Peter B. "The Purposing of High-Performing Systems." *Organizational Dynamics* (Autumn 1982).

Verimb, Carlos. "Automatic Query Adjustment in Document Retrieval." *Information Processing and Management* 13(6) (1977): 339-353.

Vickery, A. and H. M. Brooks. "Plexus-The Expert System for Retrieval." *Information Processing and Management* 23(2) (1987): 99-117.

Vogt, Eric Edwards. "The Nature of Work in 2010: Convergence and the Workplace." in *Annual Review of Institute for Information Studies.* Falls Church, VA: The Institute, 1995.

Wall, Larry, Tom Christiansen, and Randal L. Schwartz. *Programming Perl*, 2d ed. Sebastopol, CA: O'Reilly & Associates, Inc., 1996.

Wang, Peiling. "A Cognitive Model of Document Selection of Real Users of Information Retrieval Systems." Ph.D. diss., University of Maryland/CLIS, 1994.

Wasserman, Paul and Mary Lee Bundi, eds. *Reader in Library Administration.* Englewood, CO: Information Handling Services, 1968.

Weingand, D. E. "Grounded Theory and Qualitative Methodology." *International Federation of Library Associations Journal* 19(1) (1993): 17-26.

White, Marilyn Domas. "The Dimensions of the Reference Interview." *Reference Quarterly* (Summer 1981): 373-380.

________. "The Reference Encounter Model." *Drexel Library Quarterly* 19 (Spring 1983): 38-55.

________. "Evaluation of the Reference Interview." *Reference Quarterly* 25(1) (1985): 76-84.

Wilson, Patrick. “Situational Relevance.” *Information Storage and Retrieval* 9(8) (1973): 457-471.

Wilson, Patrick. “The Face Value Rule in Reference Work.” *Reference Quarterly* 25 (Summer 1986): 468-75.

Winograd, Terry. “Case Grammar.” in *Language as a Cognitive Process. Syntax. Vol. 1*. Reading, MA: Addison-Wesley, Inc., 1983.

Wong, S.K.M. and Y.Y. Yao. “A Probabilistic Method for Computing Term-by-Term Relationships.” *Journal of American Society of Information Science* 44(8) (1993): 431-439.

Wynar, Bohdan S. *Introduction to Cataloging and Classification,* 8th ed., rev. Arlene G. Taylor. Englewood Co.: Libraries Unlimited, Inc., 1991.

Yan, W. Tak and Annevelink, Jurgen. “Integrating a Structured-Text Retrieval System with an Object-Oriented Database System.” in *Proceedings of VLDB'94* (1994).

Yan, Tak. and Garcia-Molina, Hector. *Index Structures for Selective Dissemination of Information. STAN-CS-92-1454*, 1993.

Yan, Tak. and Garcia-Molina, Hector. *Index Structures for Information Filtering Under the Vector Space Model. STAN-CS-93-1494,* 1993.

Yan, Tak. and Garcia-Molina, Hector. “SIFT - a tool for wide-area information dissemination,” in *Proceedings 1995 USENIX Technical Conference*, 177-186.

Yang, Yiming and Christopher G. Chute. “An Example-Based Mapping Method for Text Categorization and Retrieval.” *ACM Transactions on Information Systems* 12(3) (1994): 252-277.

Young, Degl and Ben Schneiderman. “A Graphical Filter/Flow Representation of Boolean Queries: A Prototype Implementation and Evaluation.” *Journal of American Society of Information Science* 44(6) (1993): 327-339.