

UNIVERSITY OF CALIFORNIA SANTA BARBARA

**The Organization of
Human Knowledge**
Systems for Interdisciplinary Research

A Thesis submitted in partial satisfaction of the
requirements for the degree Masters of Science in
Media Art and Technology by

Rama C. Hoetzlein

Thesis Committee:

Professor George Legrady, Chair

Professor Matthew Turk

Professor Alan Liu

June 2007

This thesis of Rama C. Hoetzlein is approved:

George Legrady, Committee Chair

Matthew Turk

Alan Liu

June 2007

**The Organization of Human Knowledge:
Systems for Interdisciplinary Research**

Copyright © 2007

Rama C. Hoetzlein

ACKNOWLEDGEMENTS

Many thanks to my close friends

Marty White,
Adrianna Hirtler,
Mark Zifchock,
Claire Hope Fox,
Laura Dobeck,
Abram Connelly,
Scott McCasland,
Haru Ji,
the Wisdom's Goldenrod
Philosophical Foundation ¹,
and my family,

in whose company many discussions
helped to bring this work to light.

¹ <http://www.wisdomsgoldenrod.org/>

THESIS ABSTRACT

The Organization of Human Knowledge:
Systems for Interdisciplinary Research

by

Rama C. Hoetzlein

June 2007

Knowledge organization is the problem of arranging and classifying what we collectively know as a society in ways that can be easily understood and communicated to others. The issues addressed in this thesis include the representation and storage of knowledge, semiotics, ontology, classification, systems for knowledge organization, and the visualization and aesthetics of knowledge systems. The Quanta software system is presented as a generic framework for the integrated storage, organization and visualization of human knowledge in interdisciplinary contexts.

Novel contributions are made to both technical and conceptual aspects of knowledge organization. Technical contributions include a hypergraph structure for the storage and efficient representation of knowledge, comparative zoomable timelines for the visualization of events in time, circle packing with dynamic loading to visualize trees, and a distributed architecture and protocol for social knowledge systems. Conceptual contributions include a new measure of meaning in data systems, the data-semantic ratio, an analysis of the relationship between the semiotic triangle and the data-information-knowledge triangle, and motivations for knowledge visualization as a field of study.

Topics on the philosophical, social and technical aspects of knowledge organization are considered in historical context with an emphasis on interdisciplinarity to provide a framework for a synthesis of ideas across various disciplines.

TABLE OF CONTENTS

Acknowledgements	iv
Abstract.....	v
Table of Contents	vi
List of Figures	xi
List of Tables	xvi
Preface	xviii
I. Introduction	1
1.1. Overview of the Thesis.....	1
1.2. Novel Contributions	8
1.3. Types of Knowledge	12
1.4. Libraries.....	15
1.5. The Internet.....	18
1.6. Integration.....	20
II. Background and Context.....	23
2.1. Social and Scientific Knowledge Organization	24
2.2. The Aesthetics of Arrangement	33
2.3. Computational Knowledge Organization	36
2.3.1. Early Database Systems.....	38
2.3.2. Scientific Visualization.....	40
2.3.3. Information Visualization	43

2.3.4. Information Aesthetics.....	47
2.3.5. Knowledge Visualization and Design	52
III. Integrative Strategies	57
3.1. Interdisciplinarity	58
3.2. Knowledge vs. Information	60
3.3. Semiotics	65
3.4. Research as Process.....	69
IV. Databases and Systems	74
4.1. Ideal Systems	75
4.2. Existing Systems	77
4.2.1. Libraries	77
4.2.2. Encyclopediae	80
4.2.3. The Internet	84
4.2.4. The Semantic Web.....	93
4.2.5. Databases	99
4.2.6. Object-Oriented Databases (OODB)	104
4.2.7. Semantic Networks.....	107
4.3. Summary and Features of Knowledge Systems.....	111
4.4. Quanta: An Interdisciplinary Knowledge Database.....	115
V. Language and Representation	119
5.1. Grammar	119
5.2. Emergent Objects	133

5.3. Patterns	136
5.4. Grammatic Layers	139
5.5. Logical Inference	143
5.6. Storage & Performance	149
5.7. Representational Layers	153
VI. Ontology and Classification.....	155
6.1. Ontology and Philosophy	155
6.2. Modern Ontology.....	163
6.3. Existential (Primary) Classification	171
6.4. Dependent Classification.....	173
6.5. Multiple Inheritance	176
6.6. Continuous Properties	179
6.7. Formal Ontology and Logic	180
6.8. Disambiguation	183
6.9. Quanta: Ontology Design	185
VII. Knowledge Visualization	189
7.1. Information Visualization and Design	191
7.2. Quantitative and Qualitative Data.....	194
7.3. Trees	201
7.4. Timelines	209
7.5. Networks.....	220
7.6. Knowledge Visualization and Cognition.....	226

VIII. Distributed Knowledge Systems	233
8.1. Background: IP, HTTP and the Semantic Web	233
8.2. SDP - A Semantic Data Protocol	235
8.3. IP Addresses and Universal Identifiers	240
8.4. Reliability and Robustness	243
8.5. Uniqueness.....	244
8.6. Order and Retrieval	245
8.7. SDP - Name Servers.....	248
8.8. A Semantic Internet Browser.....	250
8.9. Implementation	251
8.10. Social Aspects of Content.....	253
8.11. A Global Semantic Internet	258
IX. Quanta Prototype & Future Directions.....	260
9.1. Implementation Details	260
9.2. Interaction	267
9.3. Current Limitations	269
9.4. Qubit	272
9.5. Quanta Summary	276
9.6. Future Directions	276

Appendices.....	279
Appendix A. Quanta and Predicate Logic.....	279
Appendix B. Quanta Formal Ontology.....	282
Appendix C. Quanta Generic Ontology (GO).....	284
Appendix D. Qubit: Data Mining Results.....	286
Bibliography	288

LIST OF FIGURES

Figure 1.1. Quanta Conceptual Overview	4
Figure 1.2. Quanta System Overview	7
Figure 1.3. Plato's Divided Line.....	12
Figure 2.1. Tree of Knowledge, Prophyry.....	23
Figure 2.2. Systema Naturae, Carolus Linnaeus, 1735	26
Figure 2.3. Pre-Copernican view of the solar system	27
Figure 2.4. Concept map of a plant	30
Figure 2.5. John Dalton, New System of Chemical Philosophy, 1808	31
Figure 2.6. Sketches of the periodic table, Dmitri Mendeleev, 1869.....	32
Figure 2.7. O'Clock, Arman, 1998	34
Figure 2.8. Matrix of Knowledge, Agnes Denes	35
Figure 2.9. Visualization of a Storm Cloud, Tufte, 1983	41
Figure 2.10. Src Protein, David Goodsell, Scripps Research Institute.....	42
Figure 2.11. H3, Hyperbolic network layout, Tamara Munzner	44
Figure 2.12. CellViz, visualization of protein levels, Gene Network Sciences.....	45
Figure 2.13. The Visual Thesarus, Thinkmap, 1998.....	46
Figure 2.14. TextArc, Bradford Paley, 2002	47
Figure 2.15. Valence, Benjamin Fry, 2000.....	49
Figure 2.16. Seattle Library Visualization, George Legrady, 2005	50
Figure 2.17. Studyscape, Patrick Vuarnoz, 2005.....	51

Figure 2.18. Continuum of scientific, information and knowledge visualization	52
Figure 3.1. The Astronaut's Experiment	62
Figure 3.2. Parallels between the Semiotic and Information Triangles	66
Figure 3.3. Summarizing a field of study	70
Figure 3.4. Further elaboration of an arbitrary discipline.....	71
Figure 4.1. <i>Naturalis Historiae</i> , Pliny, 77 CE	80
Figure 4.2. Packets take alternate routes to complete a message.....	80
Figure 4.3. First four nodes of the ARPANet, 1969	86
Figure 4.4. Operation of a search engine.....	88
Figure 4.5. Hierarchical database model	100
Figure 4.6. Network and Relational Database models	101
Figure 4.7. Example of an Object-Oriented Database in C++.....	105
Figure 4.8. Semantic Network	107
Figure 4.9. Summary of language, communication and knowledge systems throughout history	112
Figure 4.10. Quanta: A knowledge framework supporting interdisciplinary collaboration.....	115
Figure 4.11. Quanta: Internal structures and layers	116
Figure 5.1. Semantic networks for a set of statements about Turing	122
Figure 5.2. Hypergraph networks for statements about Turing	124
Figure 5.3. Two statements represented as a generative grammars and as a hypergraph network	127

Figure 5.4. Hypergraph networks with algorithms to represent two paintings by Leonardo da Vinci	130
Figure 5.5. Complexity in databases and formal languages	132
Figure 5.6. Generative grammars used to parse a simple sentence	134
Figure 5.7. Semantic hypergraph for statements about three composers	137
Figure 5.8. Patterns of representation at different scales.....	153
Figure 6.1. Objects with different properties.....	156
Figure 6.2. Different species of objects with different properties	156
Figure 6.3. An object as something changing in time.....	157
Figure 6.4. Objects composed of more fundamental material.....	158
Figure 6.5. Six objects classified by shape, fill style and shading style.....	166
Figure 6.6. Explicit specification of taxonomies for six objects.....	168
Figure 6.7. Multiple taxonomies possible for a single quality.....	169
Figure 6.8. Circle and square combine to form an ellipsoid.....	173
Figure 6.9. New properties can arise in inherited classes.....	177
Figure 6.10. A circle becomes a square.....	179
Figure 6.11. Existential and classification hierarchies in the Quanta Ontology.....	186
Figure 7.1. Celestial Atlas, A. Jamieson, 1820.....	193
Figure 7.2. Atomic number versus melting point for the Elements	195
Figure 7.3. Graph visualization: Atomic mass and density of the Elements.....	197
Figure 7.4. Graph visualization: Atomic group and hardness of the Elements	197
Figure 7.5. Graph visualization: Heat capacity and speed of sound	198

Figure 7.6. Graph visualization: Earth and solar abundance for the Elements.....	198
Figure 7.7. Indented list, with thumbnail icons, in Quanta	201
Figure 7.8. Circle packing algorithm.....	204
Figure 7.9. Circle packing: View of the Linnean taxonomy of organisms	206
Figure 7.10. Circle packing: View of mineral compounds with images	207
Figure 7.11. Chart of a Biography, Joseph Prestley, 1765	210
Figure 7.12. TimeQuilts, Huynh et al., 2005	211
Figure 7.13. Timeline collision algorithm	213
Figure 7.14. Timeline visualization: Modern Sculpture, Zoom 1700 to 2200	215
Figure 7.15. Timeline visualization: Modern Sculpture, Zoom 1930 to 2010	216
Figure 7.16. Timeline density plot of Painters and Sculptors.....	217
Figure 7.17. Comparative Timeline of Sculptures and Papers on the topic of Physically-Based Modeling (subfield of Computer Graphics)	219
Figure 7.18 Network visualization on nested spheres.....	223
Figure 7.19. Analogies between human cognition and visualization systems	229
Figure 8.1. Relationship of standard Internet protocols	236
Figure 8.2. Partitioning a hypergraph across multiple machine	237
Figure 8.3. Example hypergraph stored on multiple machines	240
Figure 8.4. Autonomous information storage in a distributed system using a fluid analogy.....	242
Figure 8.5. SDP Name servers	249
Figure 8.6. Authoritative management of a distributed ontology.....	254

Figure 8.7. Filter-based management of a distributed ontology	255
Figure 9.1. First test of the core Quanta database, 2004	261
Figure 9.2 Sample input data in an Excel spreadsheet.....	263
Figure 9.3. Screenshot of Qubit downloading paintings from Wikipedia.....	274
Figure 9.4. Screenshot of Quanta after minerals have been downloaded from Wikipedia using Qubit	275

LIST OF TABLES

Table 2.1. Classification of fish, bears and humans according to two schema	29
Table 3.1. Definitions of data, information and knowledge with examples.....	64
Table 3.2. Summary of definitions of data, information and knowledge.....	68
Table 3.3. Taxonomies for the fields of music and computer graphics	70
Table 4.1. Chapter organization in Pliny's <i>Naturalis Historiae</i> , 77 CE.....	81
Table 4.2. Google search on the keyword "music"	90
Table 4.3. Google search on the phrase "robot that draws"	92
Table 4.4. Different uses of XML as metadata.....	95
Table 4.5. Desirable features of a knowledge system	113
Table 4.6. Quanta: Description of functional layers	118
Table 5.1. Translating a relational database back into English language	120
Table 5.2. Quanta: Grammatical layers	141
Table 5.3. Hypergraph model of nine sentences with full indexing and dictionary compression.....	148
Table 5.4. Quanta: Data storage layers and algorithms.....	151
Table 6.1. Ontologies according to Aristotle and Porphyry	160
Table 6.2. Two dimensions of reality: physicality and actuality	161
Table 6.3. Quanta base ontology for man-made entities	175
Table 6.4. Quanta: Ontology Design Principles	187
Table 6.5. Quanta: Ontological Layers.....	188

Table 7.1. Correlation between growth in data nodes for the Quanta ontology and unit sphere surface areas	225
Table 7.2. Scientific, Information and Knowledge Visualization.....	231
Table 8.1. User query on the word "house" using a non-authoritative organization of the semantic internet	256
Table 9.1. Visualizations in Quanta	264
Table 9.2. Data sets in the Quanta prototype, as of June 2007	266

Preface

Before the invention of television, radio and the airplane - less than two hundred years ago - people existed in a world in which communication required days to reach distant lands. In such a world, it is possible for a person to be born into a specific environment with local customs and beliefs so that there is little need to confront far-reaching ideas and new systems of thought. Consider the following statement:

Life is a mystery.

Imagine existing in a time where daily changes occur, where the work of a lifetime unfolds but collective belief in this deeper truth remains constant. Where the negotiations of trade and business occur with others who hold the same world view for they were born into the same structures. Such a person might find it easy to continue living at relative peace with the world despite life's ongoing cycles. Now consider the following:

Life is a process of evolution.

Life is a divine creation.

Life describes anything capable of reproduction.

Life is a coded molecular sequence.

Life is present in all things.

While debates on the nature of life continue, perhaps indefinitely, the problem confronting the modern individual is that we now receive not just one but *all* of these messages simultaneously and at an early age. The traditional communication of fundamental belief, normally passed down through parental guidance and story telling and which typically conveys a few fundamental yet consistent truths has been replaced by media, television and the Internet which convey many possibly conflicting ones together.

This creates a double-edged sword. On the one hand, it generates a great deal of confusion about who we are. We have entered the equivalent of Paul Virilio's *Information Bomb* in which such a great mass of ideas now confronts us that the individual feels lost. Yet the symptoms may all be related to a deeper contemporary problem: Lacking a common belief system, on what set of principles do we resolve philosophical, religious and scientific differences? How do six billion people communicate?

On the positive side, we also find ourselves with far greater access to many different ways of thinking. Advances in communication and travel allow us near immediate access to beliefs of the entire world while science develops an empirical framework for fundamental thought. Yet even science has splintered into relative truths with the formation of disciplines. Through these disciplines, fields of study may go into greater depth but conclusions that

synthesize many areas simultaneously are increasingly rare. Science, from the point of view of the individual, has become a web of distinct fields such that it is increasingly difficult to distinguish the minor truths from the major ones.

The organization of knowledge is not, therefore, simply the movement of information to proper places. Rather knowledge organization is the fundamental interdisciplinary problem of placing concepts and ideas into a context within the whole. Unlike the Internet, which is still relatively disorganized, presentations of the world's ideas should be readable, understandable and clarifying. There is no one unique solution, but we can make what we know easier to see and comprehend. In this respect knowledge organization is more similar to comparative philosophy than to information management. The goal of knowledge organization is to form bridges across disciplines, peoples and places. It is to arrange what is and can be known on a relative scale.

The task should not be seen as a rigid one but a dynamic, flexible process whose goal is to construct a new individual identity of ourselves in which shifting beliefs are placed in a proper context with respect to one another.

Rama C. Hoetzlein

Santa Barbara, CA
Dec 15, 2005

