

AUTOMATING KNOWLEDGE ACQUISITION FROM NATURAL LANGUAGE TEXT

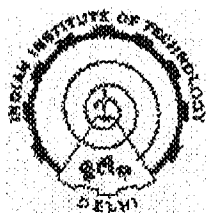
BY

RAJESH BHAT

DEPARTMENT OF MATHEMATICS

Submitted

in fulfillment of the requirements of the degree of Doctor of Philosophy
to the

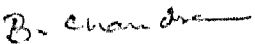


Indian Institute of Technology, Delhi
May & 1998

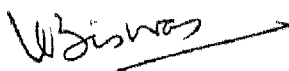
CERTIFICATE

This is to certify that the thesis entitled **Automating Knowledge Acquisition from Natural Language Text** which is being submitted by **Rajesh Bhat** for the award of **Doctor of Philosophy** to the **Indian Institute of Technology, Delhi** is a bonafide record of research work under our guidance and supervision.

The thesis has reached the standard of fulfilling the requirements of the regulations relating to the degree. The results obtained in this thesis have not been submitted to any University or Institute for the award of any degree or diploma.


Dr.(Mrs.) B. Chandra
Professor

Computer Applications Group
Department of Mathematics
Indian Institute of Technology
New Delhi


Dr. K. K. Biswas
Professor

Artificial Intelligence and Robotics
Department of Computer Science and Engineering
Indian Institute of Technology
New Delhi

Acknowledgements

With great pleasure and deep gratitude, I acknowledge the invaluable guidance of Prof.(Mrs.) B. Chandra and Prof. K. K. Biswas, my thesis supervisors. In spite of their busy schedule, they could find time to provide precious guidance. They inspired me to cultivate new research ideas. I am extremely grateful to Prof.(Mrs.) B. Chandra for her constant suggestions and discussions throughout my research period. But for her help, this success would not have been possible. I express my sincere gratitude to both the supervisors and I am deeply indebted to them.

I also express my sense of indebtedness to the Dean P.G. Studies, Head of Mathematics Department and other faculty members of the Department for the support they have provided for pursuing my research. I also give due credit to the Head of Computer Center, my officemates past and present for providing the conducive environment and moral support.

Special Thanks to Prof. B. T. Kaul, University of Delhi, who stood by me through difficult times during this period. Lastly, I thank my family members and friends for their emotional support and constant encouragement for this pursuit.

Rajen Patel

Abstract

Most of the Artificial Intelligent Systems today are dependent on Domain Specific Knowledge. This dependence on Domain Knowledge requires lot of information to be already present in systems i.e. systems need pre-existence of abundant domain specific knowledge. This limitation of domain dependent systems, however, ensures practically success only in limited domains. Thus there is a need to look for a methodology which helps in development of systems capable of dealing with domain independence of knowledge. The type of source of knowledge to systems does play an important role in this direction. Natural Language Text seems to be more easily accessible source for such a purpose.

Main objective in this thesis is to develop a Domain Independent Methodology to build Artificial Intelligent Systems [AIS] which are capable of self-learning, self-organizing and self-validating knowledge. Knowledge plays an important role in building Intelligent Systems. Two important aspects in Knowledge Engineering namely Knowledge Representation and Knowledge Acquisition have been dealt in this thesis. The structure [**< Object> - Relation⁺ - <Object⁺>**] (**ORO**) has been developed as a knowledge representation technique to represent knowledge extracted from sentences in a text describing an application domain.

Three different system layers constituting an Artificial Intelligent System (AIS) based on **ORO** constructs, have been defined. First Layer termed as **Natural Layer** represents Natural Language Text expressing knowledge inherent in an application domain. Knowledge in Second

Layer i.e. in **Cortical Layer**, is extracted from Natural Language Text using **Adaptive Resonance Theory** model. Third Layer i.e. **Intelligent Layer** contains a refined Knowledge Base. This structure of **ORO** has further been refined to [**<Object-class> Relation⁺ <Object-class⁺>**] within framework of **Activity Structure**. Topologies on an AIS model, based on Activity Structure concept, have been defined in the form of a large number of rules which help in automating knowledge acquisition by self-learning and self-organizing knowledge in Intelligent Layer. Initially, an example text is used for acquiring knowledge in an Intelligent Layer (**Core**) of a system. Considering Knowledge-Base free from any context of such an example text, provides a universal language-free conceptual reasoning system for understanding text of any other domain or a similar domain in future. The redeeming feature of the methodology is that **future** learning of knowledge of any domain is without interacting much with background knowledge which otherwise requires lengthy lexicon of language to be pre-codified and pre-existing in an AIS.

Table of Contents

CHAPTER 1

Introduction	1
1.1 Past Work on Knowledge Representation and Acquisition	4
1.2 Thesis Overview	12

CHAPTER 2

Structured Knowledge Representation- Object Relation Object Construct	18
2.1 Introduction	18
2.2 Object-Relation-Object Construct	21
2.3 AIS in an Application Domain	25
2.4 Case Studies	32
2.5 Comparative Study of ORO	42
2.6 Algorithm	59

CHAPTER 3

Classification and Semantics of Objects and Relations	62
3.1 Introduction	62
3.2 System layers	64
3.3 Classification	66
3.4 Semantics of Objects and Relations	72
3.5 Interactive Pragmatics Learning	78

CHAPTER 4

Adaptive Resonance Theory Model for Generation of ORO Construct	90
4.1 Introduction	90
4.2 ART1 Architecture	92
4.3 ART Model for The Generation of ORO Constructs	96
4.4 Design of The ART Model	100
4.5 Algorithms	102
4.5.1 Algorithm 1	104
4.5.2 Algorithm 2	105
4.5.3 Algorithm 3	107

CHAPTER 5

Refinement of ORO Construct within The Framework of Activity Structure	109
5.1 Introduction	109
5.2 A Brief Sketch of Activity Structure	110
5.3 Design of An Information Processing Machine(IPM)/ Knowledge- Based System by Means of Activity Structures	116
5.4 Mathematical Modelling of Conceptual Realization of AIS	124
5.5 The Refinement	127

CHAPTER 6

Topologies on The Activity Structure Model of An AIS	131
6.1 Introduction	131
6.2 Basic Operations on Objects and Relations	132
6.2.1 Relations	133
6.2.2 Reverse Synonym Relations	134
6.2.3 Create Relation	135
6.2.4 Synonym Relations	137
6.2.5 Group Relations into Relation-Class	140
6.2.6 Synonym/Merge Objects/Object-Classes	141
6.2.7 Group Objects/Object-Classes	145
6.2.8 Create Object/Object-Class	147
6.2.9 Unidirectional Relation	149
6.2.10 Destroy Relation	149
6.3 Analysis	150
6.3.1 Different Ways of Analysis	151
6.4 Examples	154
6.4.1 Example 1	156
6.4.2 Example 2	159
6.4.3 Example 3	162
6.4.4 Example 4	167
6.5. Conclusions	171
References	176
Appendix	181
Glossary of Terms	19