Chapter 1: Introduction

1.1. Motivation

1.1.1 Causes, Symptoms and Treatment of Parkinson's Disease

Parkinson's Disease (PD) is a degenerative disease that affects the Central Nervous System (CNS) due to the premature death of collection of cells in the brain stem. The first report on PD was described by James Parkinson in 1817, based on his observations of six individuals, although the ancient Indian medical treatment of Ayurvedic medicine describes a disorder similar to PD [1]. The depletion of these cells leads to the reduction of the chemical known as "dopamine" which controls the neuromuscular movement of the body. The chemical structure consists of a benzene ring, two hydroxy groups and an amino group as shown in Figure 1.1.

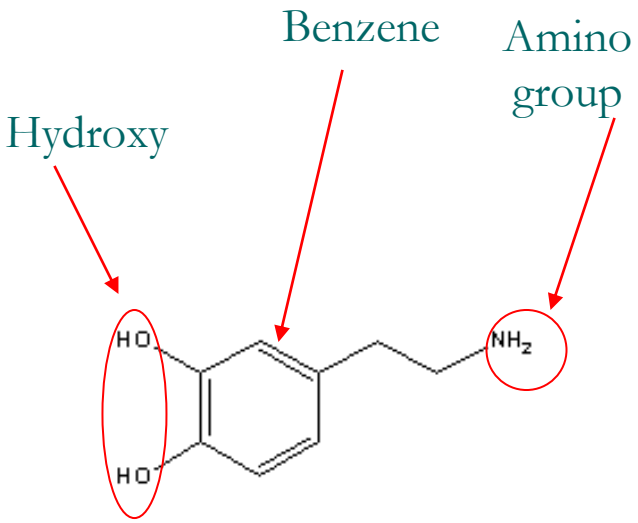

Figure 1.1. The chemical structure of dopamine

Dopamine is a neurotransmitter that carries messages across the synapse and allows the **striatum** to both initiate and control the movement and balance of the body [1]. It

ensures the smooth, voluntary movement of the muscles. Dopamine messages are packaged in packets by the nerve cells that transmit them across the synapse into the neighbouring nerve cell, which in turn passes the message to the next nerve cell. This is shown in Figure 1.2.

Once the packets are received in the next nerve cell, they then release the dopamine particles back into the synapse. They then fine-tune the coordination of the movement between the used dopamine particles along with any excess dopamine that did not initially fit into the nerve cell are broken down in the synapse by the chemical known as MAOB [1]. It is this depletion or excess in dopamine that causes PD.

The 1997 Charter for People With Parkinson's Disease (PWPD) states that a patient has the right to consult a doctor with special interest in Parkinson's Disease and expect the following [5]: (a) receive an accurate diagnosis, (b) have access to support services, (c) receive continuous care and (d) take part in managing the illness. In industrialised countries, the prevalence of the disease is "estimated at 0.3% of the general population and about 1% of the population older that age 60 years"[12] and affects all sections of ethnic groups, with men being more prone to the disease [13]. Although the disease affects around 100,000 people alone in the United Kingdom, there are no biochemical tests to diagnose the disease [4] to date. Kale et al [5] cites the usage of a multi-disciplinary team like doctors, specialist nurses, speech therapists, physiotherapists and dietitians to provide for the care for PWPD. While clinical diagnosis is based on the experience of neurologists, geriatricians and general practitioners [5], modern imaging techniques like the PET SCAN or single photon computed emission tomography (SPECT)

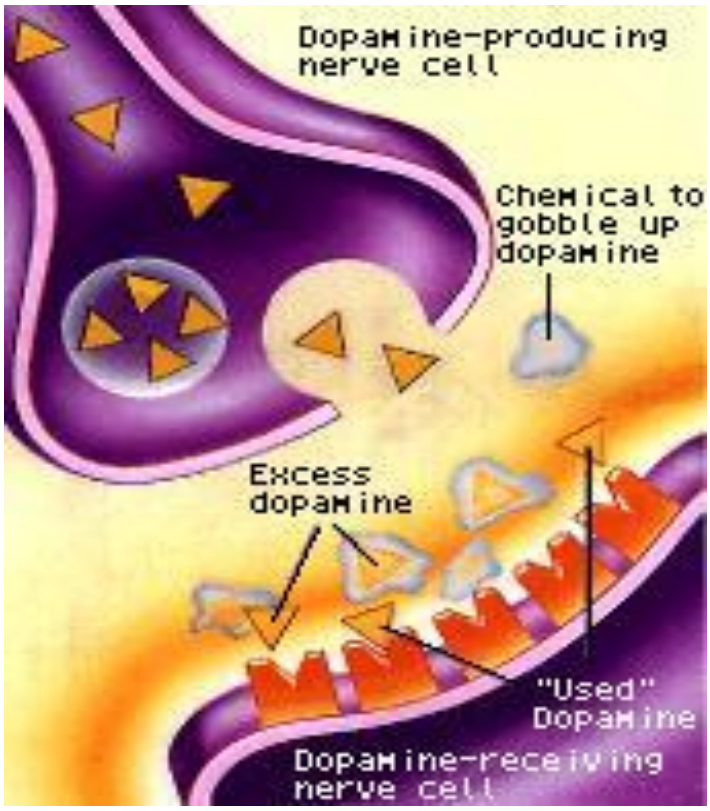

Figure 1.2.Dopamine is transferred across the synapse [1]

and, more recently, DaTSCAN can aid in the diagnosis of therapeutic trials. The clinical symptoms of PD are as follows:

- resting tremor on one side of the body
- slowness of movement (**bradykinesia**)
- stiffness of limbs (**rigidity**)
- gait problems
- small cramped handwriting (**micrographia**)
- lowered voiced volume (**dysarthia**)
- feelings of depression
- increase in dandruff or oily skin
- less frequent blinking and swallowing

Studies have been conducted to find the known causes for PD like pesticides [6], slowness of the dietary supplementation of the naturally occurring coenzyme Q_{10} [7] and even the mutation of genes [8] in order to help expedite the cure for the disease.

The treatment of PD began with the usage of levodopa in the early sixties [9] but can give rise to motor fluctuations and **dyskinesias** (i.e. impairment of voluntary movements resulting in fragmented or jerky motions), as long term consequences of using the medicine [10]. The wearing down effect of using levodopa is noticeable within a few years. Hence the goal of effective pharmacological treatment lies in the achievement of satisfactory symptomatic benefits while minimizing the side effects [11]. As such, the treatment of PD is carried out with the effective management of pharmacological drugs along with other non-pharmacological measures such as dietary considerations, speech therapy and exercise. The effective management of the drugs is essential since the effect of the drugs diminish with the passage of time, also known as the 'on-off'effect [2].

1.1.2 Software Integration of Neurological data sources

It is necessary to collate all these information sources so that medical practitioners, health care givers, scientists and researchers can further understand the workings of this incurable disease that will help to contribute in the medical developments in the future. The grand goal of neuroinformatics lies in reconciling disparate data sources and integrating data obtained from different but conceptually similar studies, sub-disciplines and laboratory systems. Some of these approaches involve the creation of data warehouses that permits the consolidation and organization of data from multiple sources in heterogenous platforms [14], while others involve the creation of "virtual federated" databases which allows the maintenance of separate databases in formats like XML that enables the content to be understood by a federated engine like mediators[15].

Ontologies can be regarded as a representation of the knowledge that specifies the concepts in a specified domain [14] and provides a framework that facilitates the interoperability of heterogeneous databases. It captures the hidden meanings of the discourse. Such metadata helps machines to interact with each other since the underlying meaning of the application domain is built in conjunction with domain experts. Ontologies form the modeling constructs in developing a knowledge base. While databases containing legacy data can be mined to draw inferred knowledge, ontologies serve as building blocks in the construction of knowledge.

Neurological database like Cell Centered database[♣] (CCDB) contains data on spiny dendrites as well as 3-D structures of cells and subcellular structures. The CoCoMac[♦] database contains neuroanatomical data like cortical activity of the marquee brain. The Senselab[♠] of Yale University integrates information on the olfactory system including the protein sequences of the system. It is this lack of standardization in the available nomenclature in databases that can cause confusion to practitioners and researchers alike. Ontologies help reduce the semantic ambiguity that is prevalent in the databases.

1.1.3 Research Purpose

The absence of any real time, clinical databases coupled with the lack of standardisation of available data was the driving force behind the project. It was necessary that the project incorporated clinical and therapeutical databases that can then be integrated into the different knowledge bases. The purpose of this research is therefore to investigate the design and implementation of knowledge base systems for PD that will tap into the intellectual resources of domain experts and help in the retrieval of meaningful information. In achieving this purpose, we developed a framework that provided

- modeling of DAML+OIL enabled knowledge base using a new UML Profile
- implementation of the ontologies using the process of a knowledge map, manual and semi-automatic mapping of knowledge bases
- validation of knowledge bases to reduce the knowledge gap
- design and implementation of new databases
- integration of databases into ontologies to reduce the semantic heterogeneity of databases

[♣] http://ccdb.ucsd.edu/CCDB/index.shtml

[♦] http://cocomac.org

[♠] http://senselab.med.yale.edu/senselab

1.1.4 Research Questions

This research tries to address the following topics:

(1) How can we model the knowledge base system? What tools exist in the software industry that will allow the modeling of both the information and the knowledge base system?

This has been made possible by the use of UML. UML is an industry standard that has been used in the modeling of information system. Using the lightweight extensions of UML Profile, a DAML+OIL enabled knowledge base has been modeled. The class diagram of UML has been used to map the concepts and properties of ontology. While it was possible to map the relationships between concepts and properties, the problem remained in modeling axioms.

(2) How can we aid both clinicians and researchers to determine whether socio-economic factors like income level of patients have any effect on the drug concordance of PD?

In this case, we used relational data model to construct a clinical and a physiotherapy database. The data is gathered in case of the clinical database by the use of a semantically enabled web. However, this raised another issue since it became essential to integrate the data sources with the knowledge base.

(3) How do we reduce the semantic heterogeneity of the data sources? Is it possible to extract ontology from the data sources?

The semantic heterogeneity of the data sources was reduced by integrating with ontology. The conceptual model of the data sources contain knowledge of the domain experts and are used in the extraction of ontology. The ontologies extracted were then enhanced by aligning with a foundational ontology.

(4) How can we construct a knowledge base and is it possible to map different knowledge bases?

Knowledge bases were constructed using the process of knowledge map. The knowledge bases were enhanced by the addition of axioms. Similarities between the knowledge bases were removed by mapping between the knowledge bases. However,

manual mapping was found to be tedious and hence a semi-automatic mapping was employed.

1.2 Layout of Thesis

This thesis explores the following topics: (a) modeling of ontology using UML, (b) the integration of heterogeneous data sources with ontology, (c) the principle of knowledge mapping for the various knowledge bases, (d) the PD clinical and physiotherapy databases. The Figures 1.3 and 1.4 show the relations between the categories that have been described in the dissertation.

A description of each of the following chapters is described below.

Chapter 3 implements the E-R data model for **P**arkinson's **D**isease **C**linical **D**ata**b**ase (PDCDB) and **P**arkinson's **D**isease **Ph**ysiotherapy **D**ata**b**ase (PDPHDB). The PDCDB is the backend for the semantic web enabled portal for the Guildford Parkinson's Disease Research Group (GPDRG). It is classified in 6 categories: (a) Activities of Daily Living, (b) Drug Concordance, (c) Stages of Parkinson's Disease, (d) Parkinson's Disease Symptoms, (e) Side Effects and (f) Services and support. Similarly, the PDPHDB is also classified into 5 categories. They are (a) Techniques, (b) Services, (c) Assessment and Carer Needs, (d) Communication and (e) Carer and Physiotherapists Information.

Chapter 4 simultaneously integrates the ontologies of heterogeneous databases. PDCDB and PDPHDB are abstracted for ontology by reverse engineering the database schema. This chapter will (a) show the export of ontology into a database, (b) reverse engineering the database schema into ontology, (c) identify ontology in multiple entities, relationships in tables, new concepts suggested from database schema and (d) ontological enhancement by aligning with Upper level ontology .

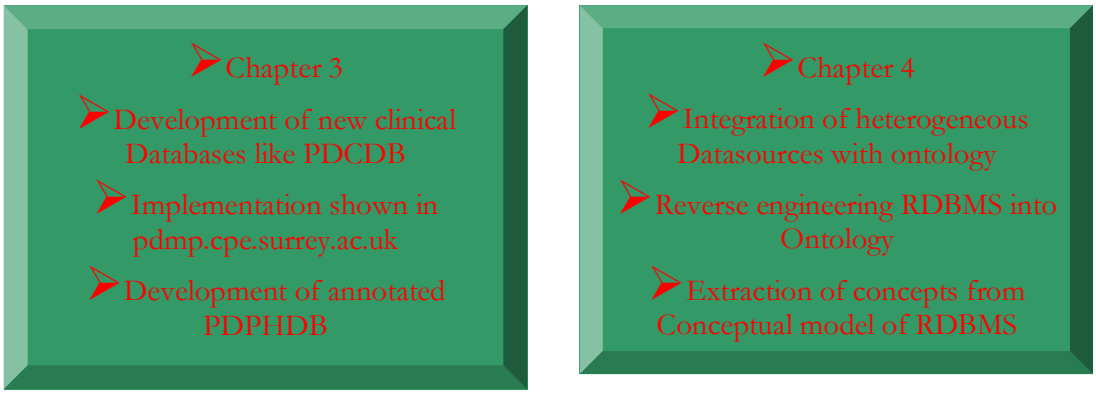

Figure 1.3: Graphical representation of chapters 3 and 4

Chapter 5 describes the modeling of knowledge engineering of the knowledge base of PDSpeechtherapy using the UML Profile. This chapter defines the mapping of UML vocabularies of packages, classes, attributes and relations to DAML+OIL ontology, concepts and properties by creating primitives for each of the knowledge bases of PD that can then be extended for any new development of knowledge concepts and attributes.

Chapter 6 provides a large scale implementation of the real world knowledge map of PWPD. It consists of the knowledge base for Symptoms and drugs (PDSymptomsdrugs), knowledge base for diet (PDDiet), knowledge base for physiotherapy (PDPhysiotherapy) and knowledge base for Speech and language therapy (PDSpeechtherapy). The inference of the knowledge is also shown. The development of the knowledge base was exemplified using PDSymptomsdrugs outlining the classification of the concepts, the relationship between attributes and the domain lexicons. The knowledge base was enhanced by the addition of axioms in the form of hyponyms, enforcing domain-specific consistency, enumeration of external services provided by physiotherapists. Finally, the mapping and merging of ontologies that show the similarities were done both manually and semi-automatically using PROMPT.

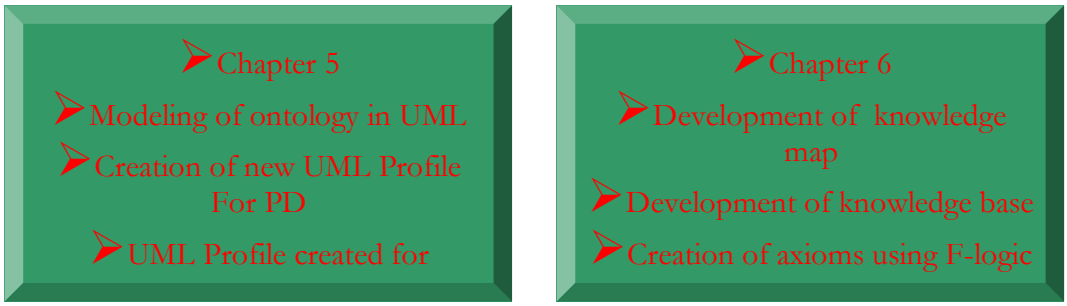

Figure 1.4: Graphical representation of chapters 5 and 6

Chapter 7 describes the conclusions reached, summarizes the achievements in the dissertation and provides recommendations for future work