



**UNIVERSITI PUTRA MALAYSIA**

**RANGKAIAN NEURAL UNTUK  
SISTEM DAPATAN SEMULA PERKATAAN  
DARIPADA PANGKALAN DATA**

**NUR IZURA UDZIR**

**FSAS 1998 6**

**RANGKAIAN NEURAL UNTUK  
SISTEM DAPATAN SEMULA PERKATAAN  
DARIPADA PANGKALAN DATA**

**Oleh**

**NUR IZURA UDZIR**

**Tesis ini dikemukakan sebagai memenuhi keperluan  
bagi mendapatkan Ijazah Master Sains di  
Fakulti Sains dan Pengajian Alam Sekitar  
Universiti Putra Malaysia**

**Julai 1998**



## DEDIKASI

*“Sesungguhnya solatku, pengorbananku, hidupku, dan matiku  
adalah kerana Allah Tuhan Sekelian Alam.”*

*Suami tercinta... Samsuddin Musa*

*...terima kasih atas pengertian, dorongan dan pengorbanan abang.*

*Bonda yang dirindui... Hjh. Rashidah Hj. Said (doaku untukmu),*

*Ayahanda yang dikasihi... Hj. Udzir Abdul Hamid,*

*Bonda yang dihormati... Hjh. Khalijah Hj. Ayob,*

*Adinda tersayang... Angah, Siti, Murni, Huda, Yop, Sobri, Yasir, Adik...*

*...juga Khilmi dan Nadhrah.*

*Ayahanda Musa dan bonda Embong,*

*abang-abang dan kakak-kakak serta adik-adik...*

*...semoga ikatan kekeluargaan ini berpanjangan dan dirahmati Allah.*

*Teman-teman...*

## PENGHARGAAN

Dengan nama Allah Yang Maha Pemurah Lagi Maha Mengasihani. Segala puji syukur bagi Allah s.w.t. Pencipta, Pemilik dan Pengatur sekalian alam. Salawat dan Salam ke atas Junjungan Besar Nabi Muhammad s.a.w.

Setinggi penghargaan dan jutaan terima kasih diucapkan kepada penyelia saya, Dr. Md. Nasir Sulaiman atas segala bimbingan, galakan, nasihat serta tunjuk ajar yang telah diberikan sepanjang kajian ini dijalankan. Penghargaan ini juga ditujukan kepada Dr. Ramlan Mahmod, Dr. Hjh. Fatimah Ahmad dan Dr. Ali Mamat atas nasihat dan bimbingan.

Saya ingin mengucapkan terima kasih kepada Jabatan Sains Komputer kerana telah menyediakan kemudahan infrastruktur untuk menjayakan kajian ini. Kepada para pegawai akademik dan bukan akademik juga diucapkan terima kasih atas bantuan, sokongan dan dorongan yang diberikan.

Penghargaan juga dirakamkan untuk Universiti Putra Malaysia kerana telah membiayai pengajian saya. Ucapan terima kasih kepada staf-staf Pusat Pengajian Siswazah.

Penghargaan istimewa buat suami saya, Samsuddin Musa yang memahami dan sentiasa memberikan semangat dan dorongan sepanjang pengajian ini. Tidak lupa juga kepada ayahanda Hj. Udzir Abdul Hamid dan keluarga di atas doa yang diiringkan.



Akhir sekali, tidak dilupakan penghargaan ini ditujukan kepada semua rakan seperjuangan yang telah banyak memberikan perangsang serta bantuan - Zura, Gee, Ummu, Razali, Ina, Shiela, Karen, Maya, Liza, Kak Saba, adik-adik Baitul 'Izzah, adik-adik KBM dan PKPIM, juga teman-teman seperjuangan Helwa ABIM. Juga kepada pihak yang terlibat sama ada secara langsung atau tidak langsung dalam menyempurnakan kajian ini.

Semoga Allah memberkati kalian. Amin Ya Rabbal 'Alamin.



## KANDUNGAN

Muka Surat

<b>PENGHARGAAN</b> .....	iii
<b>SENARAI JADUAL</b> .....	vii
<b>SENARAI RAJAH</b> .....	x
<b>SENARAI SINGKATAN</b> .....	xii
<b>ABSTRAK</b> .....	xiii
<b>ABSTRACT</b> .....	xv
<b>BAB</b>	
<b>I PENGENALAN</b> .....	1
Latar Belakang Masalah .....	2
Objektif Kajian .....	3
Skop Kajian .....	4
Penyusunan Tesis .....	5
<b>II SOROTAN LITERATUR</b> .....	7
Pangkalan Data dan Dapatan Semula Maklumat .....	7
Definisi Pangkalan Data .....	8
Pengurusan Pangkalan Data .....	10
Dapatan Semula Maklumat .....	11
Pertanyaan .....	11
Teknik-teknik Konvensional dalam Dapatan Semula Maklumat .....	14
Rangkaian Neural .....	18
Sejarah Ringkas Rangkaian Neural Buatan .....	19
Gambaran Umum Rangkaian Neural Biologi .....	22
Definisi Rangkaian Neural Buatan .....	23
Topologi dan Pemprosesan Neuron .....	24
Operasi Rangkaian Neural .....	26
Contoh Model-model Rangkaian Neural .....	28
Aplikasi Umum Rangkaian Neural .....	31
Dapatan Semula Maklumat dan Rangkaian Neural .....	32
<b>III KAEDAH PENDEKATAN KAJIAN</b> .....	38
Rangkaian CPN untuk Sistem Dapatan Semula Maklumat .....	38
Kaedah Pembangunan Sistem .....	40
Perwakilan Input .....	42
Perwakilan A .....	44



Perwakilan B .....	45
Perwakilan C .....	45
Perwakilan D .....	47
Perwakilan E .....	48
Perwakilan F .....	49
Perwakilan G .....	50
Peraksanaan Sistem .....	51
Bahagian Pra-Pemprosesan .....	52
Proses Latihan .....	53
Proses Panggil-Semula .....	60
<b>IV KEPUTUSAN DAN PERBINCANGAN .....</b>	<b>64</b>
Proses Latihan dan Pengecaman oleh CPN .....	64
Eksperimen-eksperimen Awal .....	67
Analisis Keputusan Eksperimen-eksperimen .....	68
Eksperimen Menggunakan Perwakilan A .....	69
Eksperimen Menggunakan Perwakilan B .....	72
Eksperimen Menggunakan Perwakilan C .....	75
Eksperimen Menggunakan Perwakilan D .....	77
Eksperimen Menggunakan Perwakilan E .....	79
Eksperimen Menggunakan Perwakilan F .....	82
Eksperimen Menggunakan Perwakilan G .....	84
Eksperimen Menggunakan 200 Data .....	87
Perbincangan .....	89
<b>V KESIMPULAN .....</b>	<b>106</b>
Kesimpulan .....	106
Cadangan Kajian Lanjutan .....	108
<b>BIBLIOGRAFI .....</b>	<b>111</b>
<b>LAMPIRAN</b>	
Jadual-jadual Tambahan .....	116
<b>BIODATA .....</b>	<b>126</b>

## SENARAI JADUAL

Jadual	Muka Surat
3.1 Perwakilan A bagi Setiap Aksara .....	44
3.2 Perwakilan B bagi Setiap Aksara .....	45
3.3 Perwakilan C bagi Setiap Aksara .....	47
3.4 Perwakilan D bagi Setiap Aksara .....	48
3.5 Perwakilan E bagi Setiap Aksara .....	49
3.6 Perwakilan F bagi Setiap Aksara .....	50
3.7 Perwakilan G bagi Setiap Aksara .....	51
3.8 Sepuluh Sasaran Unik untuk Sepuluh Data Input .....	55
4.1 Keputusan Eksperimen Menggunakan Perwakilan A .....	70
4.2 Keputusan Eksperimen Menggunakan Perwakilan B .....	73
4.3 Keputusan Eksperimen Menggunakan Perwakilan C .....	75
4.4 Keputusan Eksperimen Menggunakan Perwakilan D .....	77
4.5 Keputusan Eksperimen Menggunakan Perwakilan E .....	79
4.6 Keputusan Eksperimen Menggunakan Perwakilan F .....	82
4.7 Keputusan Eksperimen Menggunakan Perwakilan G .....	85
4.8 Keputusan Eksperimen Menggunakan Perwakilan C bagi Set 200 Data .....	87
4.9 Keputusan Eksperimen Menggunakan Perwakilan E bagi Set 200 Data .....	88



4.10	Keputusan Eksperimen Menggunakan Perwakilan G bagi Set 200 Data .....	88
A01	Contoh Data Input Bagi Satu Nama Mengikut Perwakilan Yang Digunakan .....	116
B01	Seratus Sasaran Unik untuk Seratus Data Input .....	118
C01	Peratus Pengecaman Model 1 Terhadap Setiap Set Ujian Mengikut Skema Perwakilan bagi Pusingan Latihan 50 .....	121
C02	Peratus Pengecaman Model 1 Terhadap Setiap Set Ujian Mengikut Skema Perwakilan bagi Pusingan Latihan 70 .....	121
C03	Peratus Pengecaman Model 1 Terhadap Setiap Set Ujian Mengikut Skema Perwakilan bagi Pusingan Latihan 100 .....	121
C04	Peratus Pengecaman Model 1 Terhadap Setiap Set Ujian Mengikut Skema Perwakilan bagi Pusingan Latihan 125 .....	122
C05	Peratus Pengecaman Model 1 Terhadap Setiap Set Ujian Mengikut Skema Perwakilan bagi Pusingan Latihan 150 .....	122
C06	Peratus Pengecaman Model 1 Terhadap Setiap Set Ujian Mengikut Skema Perwakilan bagi Pusingan Latihan 200 .....	122
C07	Peratus Pengecaman Model 2 Terhadap Setiap Set Ujian Mengikut Skema Perwakilan bagi Pusingan Latihan 50 .....	123
C08	Peratus Pengecaman Model 2 Terhadap Setiap Set Ujian Mengikut Skema Perwakilan bagi Pusingan Latihan 70 .....	123
C09	Peratus Pengecaman Model 2 Terhadap Setiap Set Ujian Mengikut Skema Perwakilan bagi Pusingan Latihan 100 .....	123
C10	Peratus Pengecaman Model 2 Terhadap Setiap Set Ujian Mengikut Skema Perwakilan bagi Pusingan Latihan 125 .....	124
C11	Peratus Pengecaman Model 2 Terhadap Setiap Set Ujian Mengikut Skema Perwakilan bagi Pusingan Latihan 150 .....	124



C12	Peratus Pengecaman Model 2 Terhadap Setiap Set Ujian Mengikut Skema Perwakilan bagi Pusingan Latihan 200 .....	124
C13	Peratus Pengecaman Rangkaian Terhadap Setiap Set Ujian Mengikut Skema Perwakilan bagi 200 Data dengan 50 Pusingan Latihan .....	125
C14	Peratus Pengecaman Rangkaian Terhadap Setiap Set Ujian Mengikut Skema Perwakilan bagi 200 Data dengan 100 Pusingan Latihan .....	125
C15	Peratus Pengecaman Rangkaian Terhadap Setiap Set Ujian Mengikut Skema Perwakilan bagi 200 Data dengan 125 Pusingan Latihan .....	125



## SENARAI RAJAH

Rajah		Muka Surat
1	Medan-medan dalam Rekod Pelajar .....	4
2	Contoh Fail Pangkalan Data Pelajar dengan Lima Medan dan Enam Rekod Pelajar .....	9
3	Struktur Sistem Pengurusan Pangkalan Data .....	12
4	Rangkaian Neural Biologi .....	23
5	Rangkaian Neural Buatan .....	25
6	Satu nod ke- <i>i</i> yang Ringkas .....	26
7	Rangkaian Perambatan-balik .....	29
8	Rangkaian Kohonen .....	30
9	Rangkaian <i>Counterpropagation</i> .....	31
10	Seni Bina Ringkas Rangkaian CPN .....	39
11	Sistem Dapatan Semula Pangkalan Data yang Dibangunkan .....	41
12	Seni Bina Rangkaian Neural untuk Pengecaman Nama daripada Pangkalan Data .....	54
13	Penyusutan Rantau Kejiranan .....	56
14	Seni Bina CPN dengan Sebahagian Pemberat Terakhir Selepas Proses Latihan .....	65
15	Graf Bilangan Pusingan Melawan Peratus Pengecaman bagi (a) Set Data Bebas Ralat dan (b) Semua Set Data Menunjukkan Prestasi Rangkaian dalam Eksperimen Menggunakan Perwakilan A .....	71
16	Graf Bilangan Pusingan Melawan Peratus Pengecaman bagi (a) Set Data Bebas Ralat dan (b) Semua Set Data Menunjukkan Prestasi Rangkaian dalam Eksperimen Menggunakan Perwakilan B .....	74

17	Graf Bilangan Pusingan Melawan Peratus Pengecaman bagi (a) Set Data Bebas Ralat dan (b) Semua Set Data Menunjukkan Prestasi Rangkaian dalam Eksperimen Menggunakan Perwakilan C .....	76
18	Graf Bilangan Pusingan Melawan Peratus Pengecaman bagi (a) Set Data Bebas Ralat dan (b) Semua Set Data Menunjukkan Prestasi Rangkaian dalam Eksperimen Menggunakan Perwakilan D .....	78
19	Graf Bilangan Pusingan Melawan Peratus Pengecaman bagi (a) Set Data Bebas Ralat dan (b) Semua Set Data Menunjukkan Prestasi Rangkaian dalam Eksperimen Menggunakan Perwakilan E .....	81
20	Graf Bilangan Pusingan Melawan Peratus Pengecaman bagi (a) Set Data Bebas Ralat dan (b) Semua Set Data Menunjukkan Prestasi Rangkaian dalam Eksperimen Menggunakan Perwakilan F .....	83
21	Graf Bilangan Pusingan Melawan Peratus Pengecaman bagi (a) Set Data Bebas Ralat dan (b) Semua Set Data Menunjukkan Prestasi Rangkaian dalam Eksperimen Menggunakan Perwakilan G .....	86
22	Graf Bilangan Pusingan Melawan Peratus Pengecaman bagi Set 200 Data Menggunakan (a) Perwakilan C, (b) Perwakilan E dan (c) Perwakilan G .....	88
23	Peratus Pengecaman Model 1 dengan 100 Data bagi Setiap Perwakilan Mengikut Bilangan Pusingan Latihan .....	90
24	Peratus Pengecaman Model 2 dengan 100 Data bagi Setiap Perwakilan Mengikut Bilangan Pusingan Latihan .....	94
25	Peratus Pengecaman Set 200 Data bagi Setiap Perwakilan Mengikut Saiz Lapisan Persaingan dan Bilangan Pusingan Latihan .....	98
26	Graf Bilangan Pusingan Melawan Masa Latihan bagi (a) Model 1 dan (b) Model 2 Mengikut Skema Perwakilan .....	102
27	Medan-medan dan Rekod-rekod dalam Fail Pelajar .....	110



## SENARAI SINGKATAN

ADALINE	Adaptive Linear Neuron
ASCII	American Standard Code for Information Interchange
BP	Backpropagation
CPN	Counterpropagation Network



## **ABSTRAK**

Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia sebagai memenuhi keperluan untuk ijazah Master Sains.

### **RANGKAIAN NEURAL UNTUK SISTEM DAPATAN SEMULA PERKATAAN DARIPADA PANGKALAN DATA**

Oleh

**NUR IZURA UDZIR**

**Julai 1998**

**Pengerusi : Md. Nasir Sulaiman, Ph. D.**

**Fakulti : Sains dan Pengajian Alam Sekitar**

Rangkaian neural buatan yang diaspirasikan oleh kecekapan otak manusia memproses maklumat digunakan dengan meluas dalam aplikasi-aplikasi yang melibatkan pengkelasan atau pemetaan corak. Kelebihan utamanya iaitu sifat ketegapannya dalam persekitaran hingar dan keupayaan untuk mengecam input yang tidak sempurna atau cacat menjadikannya alat yang sesuai digunakan untuk dapatan semula maklumat yang pantas berbanding kaedah pengkomputeran konvensional, bagi menangani cabaran dapatan semula yang lebih realistik.

Dapatan semula bersekutu menggunakan rangkaian neural adalah untuk mendapatkan semula maklumat (rekod) dengan betul daripada pangkalan data bila kekunci input yang cacat dimasukkan. Model rangkaian neural yang digunakan dalam kajian ini adalah rangkaian *Counterpropagation*, yang merupakan gabungan rangkaian Kohonen dengan algoritma pembelajaran tidak terselia dan rangkaian terselia



Grossberg, dengan sifat pengkelasan corak tanpa penyeliaan pada lapisan Kohonen menjadi bahagian paling penting bagi sistem.

Kajian memfokuskan penyelidikan kepada prestasi rangkaian khususnya ketepatan pengkelasan bila skema-skema pengkodan yang berbeza digunakan untuk mewakili input. Tujuh skema pengkodan telah diaplikasikan dalam kajian ini, dengan jumlah bit perwakilan dan asas pengkodan yang berbeza. Data-data yang digunakan untuk ujian merupakan set bebas ralat, set data dengan ralat tunggal dan set yang mempunyai ralat berganda.

Secara keseluruhannya semua eksperimen memberikan keputusan pengecaman yang baik, malah dengan setiap skema perwakilan yang digunakan, rangkaian telah berjaya mengecam dengan tepat kesemua set ujian dengan peratus pengecaman 100%, walaupun dengan bilangan unit persaingan, bilangan pusingan dan masa latihan yang tersendiri. Walau bagaimanapun, rangkaian yang mengaplikasi skema perwakilan dengan asas pengkodan tertentu menunjukkan prestasi yang lebih baik berbanding penggunaan skema tanpa asas pengkodan.

Kajian menunjukkan ketepatan pengkelasan dan kecekapan sistem dipengaruhi oleh bentuk perwakilan input yang digunakan, saiz lapisan persaingan serta tempoh pusingan latihan yang optimum.

## **ABSTRACT**

Abstract of thesis presented to the Senate of Universiti Putra Malaysia in fulfilment of the requirements for the degree of Master of Science.

### **NEURAL NETWORK FOR A WORD RETRIEVAL SYSTEM FROM A DATABASE**

By

**NUR IZURA UDZIR**

**July 1998**

**Chairman : Md. Nasir Sulaiman, Ph. D.**

**Faculty : Science and Environmental Studies**

Inspired by the capability and efficiency of the human brain in information processing, artificial neural networks were widely used in pattern classification and mapping applications. Their robustness in noisy environment and their ability to recognise incomplete and distorted inputs are their main advantages over the conventional computing methods for fast and correct retrieval.

Associative retrieval using neural network was developed for correct information retrieval from a certain database when given a distorted version of an input key. The model used in this study is Counterpropagation, a combination of the unsupervised training of the Kohonen net and the supervised Grossberg net, with the unsupervised pattern classification feature in the Kohonen layer being the most important part of the system.





The focus of the study was the performance of the net specifically on the classification accuracy when different coding schemes were applied to represent the inputs. Seven coding schemes were used in the research, with different bit-size (i.e. number of bits) and encoding base. As test sets, we used error free sets, single error data sets, and double error sets.

Overall, all experiments have produced satisfying results. With every encoding schemes, the network has successfully recognised all test sets given with 100% recognition, though with different combination of number of competitive units, training cycle and time. However, the nets using coding schemes with a certain base showed better performances over schemes without any encoding base.

The study has proved that classification accuracy and system efficiency are affected by the types of input representation, the size of the competitive layer and the optimum training cycle.



# **BAB I**

## **PENGENALAN**

Sesebuah pangkalan data tidak akan bermakna sekiranya data atau maklumat yang terkandung di dalamnya tidak dapat dicapai dan digunakan oleh pengguna. Oleh itu satu sistem yang akan menguruskannya diperlukan, terutama untuk mendapatkan maklumat. Dalam usaha membentuk sistem sedemikian, berbagai teknik telah dibangunkan umpamanya pencincangan, senarai bersambung, fail songsang dan indeks pokok-B (Pratt dan Adamski, 1991; Stubbs dan Webre, 1985).

Kecekapan sistem dapatan semula maklumat sangat penting bagi memenuhi keperluan menyelesaikan pertanyaan dalam talian dalam masa maklum balas yang singkat dan juga permasalahan pertanyaan yang tidak dinyatakan dengan sempurna. Keperluan ini telah menarik perhatian para penyelidik dari pelbagai bidang seperti sains komputer, kecerdasan buatan serta bidang-bidang fungsian lain seperti kejuruteraan, perubatan dan sebagainya untuk menjalankan penyelidikan dan seterusnya mendorong kepada penghasilan pelbagai kaedah dapatan semula maklumat baik dari segi perkakasan mahupun perisian.



## Latar Belakang Masalah

Seringkali terjadi insiden di mana kita terpaksa mendapatkan semula rekod dari pangkalan data hanya dengan berdasarkan satu kekunci yang selalunya tidak dinyatakan dengan sempurna, misalnya nama yang tidak dieja dengan betul atau nama yang tidak diberikan sepenuhnya. Dalam keadaan sebegini, kita memerlukan satu sistem dapatan semula maklumat yang mampu menangani pertanyaan yang tidak sempurna ini. Pengkomputeran digital konvensional memang baik dalam aplikasi pengiraan yang cepat dan tepat, tetapi tidak sesuai untuk aplikasi seperti mencari item tertentu dalam pangkalan data hanya berasaskan maklumat yang hingar atau tidak sempurna (Vassilas, 1990).

Permasalahan tersebut dapat diatasi dengan sistem rangkaian neural yang merupakan satu cabang kecerdasan buatan. Melalui perwakilan input yang betul ia sesuai diaplikasikan sebagai alat bagi mendapatkan semula maklumat dari pangkalan data. Carian untuk mendapatkan padanan terbaik antara dunia sebenar dengan perwakilan dalaman bagi 'dunia' tersebut adalah lebih baik berbanding dengan padanan biasa memandangkan ianya membenarkan perwakilan secara dalaman objek-objek dan perkaitan di antaranya (Char *et al.*, 1988; Sejnowski dan Rosenberg, 1987; Vassilas, 1990).

Satu cabang penting dalam rangkaian neural adalah ingatan bersekutu (Vassilas, 1990) yang biasanya digunakan untuk dapatan yang cepat dan tegap kerana sifatnya yang berkebolehan untuk pelaksanaan yang tegap dalam persekitaran hingar yang disebabkan oleh kecacatan atau ralat dalam corak kekunci input. Dengan

menspesifikasikan semua atau sebahagian daripada satu vektor input kekunci yang telah disekutukan dengan suatu data tertentu yang disimpan, data tersebut boleh didapati semula.

### **Objektif Kajian**

Kajian dijalankan bagi memenuhi dua objektif utama, iaitu:

1. membangunkan perisian bagi mendapatkan semula perkataan daripada pangkalan data menggunakan rangkaian neural. Input yang dikemukakan kepada rangkaian ini adalah perkataan yang berbentuk cacat atau rosak.
2. membuat penyelidikan dan menganalisa perbandingan pengaruh bentuk-bentuk skema perwakilan input yang berbeza ke atas ketepatan pengelasan rangkaian dan kecekapan sistem.

### **Skop Kajian**

Kajian dihadkan kepada dapatan semula terhadap satu medan iaitu medan nama dalam pangkalan data. Saiz pangkalan data yang terlibat dalam kajian adalah 100 dan 200 data yang terdiri daripada nama-nama yang dipilih berdasarkan agihan nama dalam buku Panduan Telefon 1996/97 (Selangor dan Wilayah Persekutuan) keluaran Telekom Malaysia Berhad. Ini adalah untuk mendapatkan panduan bagi pemilihan nama secara rawak yang adil dan lebih mewakili keseluruhan populasi di Malaysia.

Nama-nama yang telah dipilih itu seterusnya boleh mewakili rekod-rekod dalam pangkalan data yang terlibat dalam kajian. Sebagai contoh, katakan pangkalan data yang digunakan adalah fail yang mengandungi sebilangan rekod pelajar (Rajah 1). Setiap rekod terbina dari medan-medan Matrik, Nama, Program, Klas dan Alamat.

Fail Pelajar

Matrik	Nama	Program	Klas	Alamat
--------	------	---------	------	--------

Rajah 1 : Medan-medan dalam Fail Pelajar

Medan Matrik yang merupakan medan berjenis integer unik mengandungi nombor matrik pelajar, manakala medan Nama yang berjenis aksara dan tidak unik mengandungi nama penuh pelajar berkenaan. Medan Program menyimpan maklumat mengenai program pengajian yang sedang diikuti oleh pelajar tersebut sementara medan Klas pula merupakan klasifikasi pengajiannya. Alamat perhubungan pelajar terbabit disimpan dalam medan Alamat. Medan-medan Program dan Klas juga berjenis aksara dan medan Alamat pula berjenis abjad angka.

Memandangkan rangkaian neural mampu menangani masalah input yang cacat atau tak sempurna, tumpuan diberikan kepada medan Nama sebagai kekunci input untuk dapatan semula dalam kajian ini, kerana ia lebih terdedah kepada kemungkinan kesalahan ejaan, seperti nama yang tidak dieja dengan betul. Ini disebabkan tidak ada sebarang peraturan tertentu dalam penulisan sesuatu nama sebagaimana dalam medan-medan lain yang perlu mematuhi peraturan atau nahu ejaan dan penulisan yang telah ditetapkan. Medan Matrik yang merupakan medan unik dan berjenis integer tidak

dipilih untuk kajian ini memandangkan carian dengan menggunakan kekunci ini lebih mudah dan cepat melalui kaedah konvensional.

Terdapat beberapa kemungkinan ralat atau kecacatan yang biasa dilakukan oleh pengguna pangkalan data. Antaranya termasuklah:

1. ralat penghapusan, di mana satu huruf dipadam daripada data berkenaan.
2. ralat penambahan yang melibatkan ditambah kepada data asal.
3. ralat penggantian, di mana satu huruf digantikan dengan satu huruf yang lain.

Walau bagaimanapun, kajian ini hanya mengambil kira ralat penggantian bagi tujuan menguji keupayaan pengecaman sistem. Di samping set data yang tidak cacat, dua bentuk kecacatan telah digunakan ke atas set-set data masing-masing, iaitu ralat tunggal dan ralat berganda.

### **Penyusunan Tesis**

Kajian yang dikemukakan dalam tesis ini meliputi pembangunan sebuah sistem dapatan semula maklumat secara bersekutu daripada pangkalan data dengan menggunakan rangkaian neural, khususnya rangkaian CPN.

Tesis ini dibahagikan kepada lima bab. Bab pertama telah memberi gambaran umum mengenai kajian dengan pernyataan masalah dan juga tujuan kajian dijalankan. Bab kedua dalam tesis ini akan menyorot perbincangan ringkas tentang pangkalan data, sistem pengurusan pangkalan data, dapatan semula maklumat, pertanyaan serta

beberapa teknik konvensional dapatan maklumat dari pangkalan data. Bab ini seterusnya pula akan memberikan ulasan berkenaan rangkaian neural buatan, topologi dan pemprosesan neuron, operasi rangkaian neural, beberapa contoh model rangkaian neural serta aplikasi-aplikasi umum rangkaian neural yang popular. Bahagian terakhir dalam bab ini akan mengimbas kajian-kajian yang pernah dijalankan bagi mendapatkan semula maklumat daripada pangkalan data dengan rangkaian neural sebagai alatnya.

Seterusnya dalam bab ketiga pula akan merangkumi perbincangan dan huraian terperinci pendekatan yang digunakan dalam pembangunan sistem rangkaian neural untuk dapatan semula maklumat secara bersekutu daripada pangkalan data. Ini meliputi beberapa skema pengekodan yang digunakan dalam eksperimen-eksperimen dalam kajian ini. Manakala hasil analisis keputusan kajian dan eksperimen-eksperimen yang telah dilaksanakan akan dibentangkan dan diperbincangkan dalam bab keempat.

Akhir sekali, bab kelima akan memberikan ringkasan serta rumusan ke atas perolehan kajian secara keseluruhannya yang telah diperbincangkan dalam tesis ini. Perbincangan tentang cadangan kerja penyelidikan selanjutnya akan menyimpulkan bab terakhir tesis ini.

## **BAB II**

### **SOROTAN LITERATUR**

Bab ini akan mengulas karya-karya yang telah diterbitkan berkaitan dengan pangkalan data umumnya dan dapatan semula maklumat dalam pangkalan data serta penerbitan-penerbitan berkenaan rangkaian neural untuk memberi gambaran umum bidang-bidang tersebut, memandangkan kedua-dua bidang ini terlibat dalam kajian.

Kajian-kajian lepas yang telah dijalankan dalam bidang dapatan semula maklumat dengan menggunakan rangkaian neural juga diperbincangkan pada akhir bab ini sebagai asas kepada kajian.

#### **Pangkalan Data dan Dapatan Semula Maklumat**

Salah satu bidang yang besar dalam sains komputer ialah pangkalan data. Sebuah pangkalan data dibina untuk menempatkan sejumlah besar maklumat-maklumat yang berkaitan supaya mudah diuruskan. Bagi tujuan tersebut, para saintis telah membangunkan sistem capaian maklumat automatik (berkomputer) sejak tahun 1940-an lagi dengan idea asal untuk membantu menguruskan bahan-bahan saintifik yang besar yang telah ditempatkan dalam sesebuah pangkalan data (Frakes, 1992).



## Definisi Pangkalan Data

Pangkalan data merupakan suatu fail yang terdiri daripada sekumpulan rekod-rekod yang mengandungi maklumat-maklumat berkenaan sesuatu organisasi tertentu. Setiap rekod pula terbina daripada koleksi medan-medan yang merupakan atribut-atribut bagi rekod terbabit (Pratt dan Adamski, 1991; Aho dan Ullman, 1979; Korth dan Silberschatz, 1991; Salton, 1989).

Dalam persekitaran pemprosesan fail, jumlah terkecil data yang boleh disimpan adalah bit. Bit-bit dihipunkan menjadi bait atau aksara yang seterusnya terkumpul membentuk medan (Pratt dan Adamski, 1991).

Setiap rekod juga merupakan satu entiti. Pangkalan data adalah satu struktur yang boleh menempatkan maklumat tentang pelbagai jenis entiti dan juga perkaitan di antara entiti-entiti (Pratt dan Adamski, 1991). Entiti adalah seperti katanama, misalnya orang, tempat atau benda. Contohnya rekod Pelajar sebagaimana ditunjukkan dalam Rajah 2. Setiap entiti mempunyai atribut atau sifat tersendiri. Sebagai contoh, atribut bagi entiti Pelajar ialah Matrik, Nama, Program, Klas, dan Alamat. Rekod mengenal pasti perkataan-perkataan teks individu, dan setiap medan mengandungi maklumat yang berkaitan dengan setiap kemasukan perkataan, seperti sebutan, peleraian kepada sukatan, dan petunjuk nahu (Salton, 1989). Atribut-atribut atau medan-medan bagi rekod bertindak sebagai kunci bagi setiap rekod yang digunakan dalam dapatan semula maklumat (Aho dan Ullman, 1979).