TAXONOMY LEARNING FROM MALAY TEXTS USING ARTIFICIAL IMMUNE SYSTEM BASED CLUSTERING

MOHD ZAKREE BIN AHMAD NAZRI

UNIVERSITI TEKNOLOGI MALAYSIA

TAXONOMY LEARNING FROM MALAY TEXTS USING ARTIFICIAL IMMUNE SYSTEM BASED CLUSTERING

MOHD ZAKREE BIN AHMAD NAZRI

A thesis submitted in fulfilment of the
requirements for the award of the degree of
Doctor of Philosophy (Computer Science)

Faculty of Computer Science and Information System
Universiti Teknologi Malaysia

APRIL 2011

DEDICATION

To almarhum ayahanda and bonda

To my wife

May this inspires our sons and daughter

Wan Nur Aqila, Muhammad and Anas

May Allah bless us all

ACKNOWLEDGEMENT

In the name of Allah, the Most Gracious and the Most Merciful. I thank to Allah for granting me strength and guidance throughout my journey to complete this study.

I could not have completed this work without the help and assistance of a large number of people and organizations.

The author has been truly blessed to have the guidance of two stellar advisors who have defined my PhD experience. My grateful thanks to Prof. Dr. Hjh. Siti Mariyam Hj. Shamsuddin for her wisdom and encouragement. Your supervisory approach has provided me the freedom to explore new ideas. I'm grateful to Prof. Madya Dr. Azuraliza Abu Bakar for her guidance and support throughout the duration of my studies. I have counted you both as mentor and friends these past years. The author would also like to thank Prof. Dr. Jon Timmis of the University of York for inspiring me to explore immune system. I am also indebted to Universiti Kebangsaan Malaysia (UKM) for funding my PhD study. Dr Pouzi Hamzah at Universiti Malaysia Terengganu and Dr. Tang Enya Kong at Universiti Sains Malaysia also deserve special thanks for their assistance in supplying the relevant tools and other research materials.

I would like to thank the Data Mining and Optimization Research Group of Universiti Kebangsaan Malaysia, particularly Prof. Dr. Abdul Razak Hamdan, Assoc. Prof. Dr. Salwani Abdullah, Tri Basuki Kurniawan, Nasser, Saif, Yahya, Salam and Hamza for providing much needed study breaks, discussions, rants, debates and most importantly as anchors for my sanity. Finally, my wife Wan Noraidah Mohammad for your intimate understanding and patience for me whilst writing up my thesis.

ABSTRACT

In taxonomy learning from texts, the extracted features that are used to describe the context of a term usually are erroneous and sparse. Various attempts to overcome data sparseness and noise have been made using clustering algorithm such as Hierarchical Agglomerative Clustering (HAC), Bisecting K-means and Guided Agglomerative Hierarchical Clustering (GAHC). However these methods suffer low recall. Therefore, the purpose of this study is to investigate the application of two hybridized artificial immune system (AIS) in taxonomy learning from Malay text and develop a Google-based Text Miner (GTM) for feature selection to reduce data sparseness. Two novel taxonomy learning algorithms have been proposed and compared with the benchmark methods (i.e., HAC, GAHC and Bisecting K-means). The first algorithm is designed through the hybridization of GAHC and Artificial Immune Network (aiNet) called GCAINT (Guided Clustering and aiNet for Taxonomy Learning). The GCAINT algorithm exploits a Hypernym Oracle (HO) to guide the hierarchical clustering process and produce better results than the benchmark methods. However, the Malay HO introduces erroneous hypernym-hyponym pairs and affects the result. Therefore, the second novel algorithm called CLOSAT (Clonal Selection Algorithm for Taxonomy Learning) is proposed by hybridizing Clonal Selection Algorithm (CLONALG) and Bisecting k-means. CLOSAT produces the best results compared to the benchmark methods and GCAINT. In order to reduce sparseness in the obtained dataset, the GTM is proposed. However, the experimental results reveal that GTM introduces too many noises into the dataset which leads to many false positives of hypernym-hyponym pairs. The effect of different combinations of affinity measurement (i.e., Hamming, Jaccard and Rand) on the performance of the developed methods was also studied. Jaccard is found better than Hamming and Rand in measuring the similarity distance between terms. In addition, the use of Particle Swarm Optimization (PSO) for automatic parameter tuning the GCAINT and CLOSAT was also proposed. Experimental results demonstrate that in most cases, PSO-tuned CLOSAT and GCAINT produce better results compared to the benchmark methods and able to reduce data sparseness and noise in the dataset.

ABSTRAK

Fitur yang diekstrak dalam pembelajaran taksonomi dari teks yang digunakan untuk menggambarkan konteks suatu perkataan lazimnya mempunyai kesalahan (hingar) dan masalah kejarangan data. Beberapa penyelidikan telah cuba mengatasi masalah kejarangan dan hingar dengan menggunakan algoritma pengelompokan seperti Pengelompokan Aglomerat Berhierarki (HAC), Pembahagi-dua K-min dan Pengelompokan Aglomerat Berhierarki Berpandu (GAHC). Walau bagaimanapun, kaedah ini mengalami masalah perolehan kembali yang rendah. Oleh itu, penyelidikan ini bertujuan untuk mengkaji penggunaan dua penghibiridan sistem imun buatan (AIS) dalam pembelajaran taksonomi dari teks Melayu dan pembangunan alat Perlombongan Teks Berasaskan Google (GTM) untuk pengekstrakan fitur bagi mengatasi masalah kejarangan data. Dua algoritma pembelajaran taksonomi dicadangkan untuk mengurangkan masalah kejarangan dan hingar dalam set data. Algoritma pertama direka dengan menghibrid GAHC dan Rangkaian Imun Buatan (aiNet) yang dinamakan GCAINT (Pengelompokan Berpandu dan aiNet untuk Pembelajaran Taksonomi). Algoritma GCAINT mengeksploitasi *Hypernym Oracle* (HO) yang memandu proses pengelompokan berhierarki untuk menghasilkan keputusan yang lebih baik berbanding kaedah lain. Namun, HO bahasa Melayu ini mengandungi perkataan sebagai hipernim atau hiponim yang salah, justru mempengaruhi kualiti taksonomi yang terbentuk. Oleh itu, kaedah kedua dicadangkan iaitu penghibridan antara Algoritma Pemilihan Klonal (CLONALG) dengan Pembahagi-dua K-min yang dinamakan CLOSAT. Keputusan CLOSAT adalah lebih baik berbanding kaedah tanda aras tersebut. Demi mengurangkan masalah kejarangan dalam set data, GTM dicadangkan. Namun, GTM menambah jumlah ralat ke dalam set data yang seterusnya mewujudkan hubungan yang salah diantara perkataan di dalam taksonomi. Pengaruh penggunaan ukuran afiniti dengan kombinasi yang berbeza (seperti Hamming, Jaccard dan Rand) terhadap prestasi kaedah cadangan turut dikaji. Jaccard didapati lebih baik berbanding Hamming dan Rand dalam mengukur afiniti diantara perkataan. Selain itu, alat penalaan parameter automatik berasaskan Pengoptimuman Partikel Secara Berkumpulan (PSO) juga dibangunkan. Keputusan kajian menunjukkan bahawa dalam kebanyakan kes, CLOSAT dan GCAINT yang ditala PSO menghasilkan keputusan yang lebih baik berbanding kaedah lain serta mengurangkan masalah kejarangan dan hingar pada set data.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

aiNet	-	Artificial Immune Network
AIRS	-	Artificial Immune Recognition System
AIS	-	Artificial Immune System
CAIT	-	Center for Artificial Intelligence Technology of UKM
DMO	-	Data Mining and Optimization Research Group of UKM
CLONALG	-	Clonal Selection Algorithm
OL	-	Ontology Learning
PSO	-	Particle Swarm Optimization
UKM	-	Universiti Kebangsaan Malaysia

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 Introduction

One of the keystones of the Semantic Web vision is to represent information in a structured form so that the computers can "understand" and subsequently solve complex problems. Ontology is one of the main components of Semantic Web which provides a common vocabulary for a specific domain of interest. It describes properties of a term or word so that machines, applications or services can effectively share information and knowledge, thus ensuring interoperability among machines.

However, the progress towards Semantic Web is slowed down due to the knowledge acquisition bottleneck. The manual ontology modeling, as described by researchers is a process that is labor-intensive, tedious, complex, time-consuming and expensive (Alesso and Smith, 2005; Gulla and Brasethvik, 2008; Yeh and Sie, 2006; Cimiano, 2006). One of the most important components of ontology is taxonomy or also known as concept hierarchy or thesauri.

As unstructured texts such as Web page or e-books are massively available, most researchers have attempted to induce taxonomies from such resources by using machine learning, statistical analysis and natural language processing (NLP). The process of automatic knowledge acquisition for taxonomy creation is called taxonomy learning. The

(semi-)automatic support in constructing taxonomy on the basis of unstructured textual resources is referred to as *taxonomy learning from texts*.

Different learning approaches and methods from a spectrum of fields such as statistical analysis, machine learning and natural language processing have been proposed for partially or completely automatic construction of taxonomies. Alexander Maedche and Steffen Staab (2000) distinguish taxonomy learning from text techniques into: i) pattern-based extraction; ii) Association rules; iii) Conceptual clustering; iv) Ontology pruning; and v) Concept learning. However, based on available resource and research scope, it can be concluded that the association rules, ontology pruning and concept learning will not be studied in this research because of the following reasons. Concept learning technique is used to update a given taxonomy while ontology pruning technique is used to elicit an ontology by using a core ontology as guidance. Association rules are used on the data mining process to discover taxonomic relations stored in databases or text by a ready-made taxonomy as background knowledge. Thus, these techniques are used when an existing ontology or taxonomy are already exists to 'assist' the learning process. Since this research do not rely on any existing taxonomy, all these three techniques are not suitable for this research. Therefore, this research will concentrate on pattern-based extraction and conceptual clustering. Moreover, the nature or 'behavior' of texts makes conceptual clustering much more appealing than other techniques. For example, a document may have either a huge amount of irrelevant noise such as spelling errors or grammatical mistakes which lead to either sparseness or noise in the extracted features.

Furthermore, (semi-)automatic acquisition of taxonomy from Malay text is conspicuous by its absence in the research literature. Common features and linguistic patterns extracted from English text corpora have not been tested on Malay text. Therefore, this research is motivated by the believe that any learning system that is to retrieve an acceptable set of results from Malay texts must adapt to the data sparsity and noise. Currently, Artificial Immune System (AIS) has been the attention to many researchers in dealing with data sparseness. AIS is inspired from the human immune system, and it is known for its robustness, reliability and adaptability. Thus, the natural immune system exhibits many properties that are of interest to this area of taxonomy learning from text.

Therefore, this thesis is concerned with the design, implementation and evaluation of two hybrid conceptual clustering methods which are based on Artificial Immune System (AIS) for learning taxonomy from Malay text. Both algorithms proposed in this thesis perform taxonomy learning tasks over sets of documents.

1.2 Background of the Problem

The realization of the Semantic Web depends on the broad availability of semantic resources, often incorporated in ontology. Taxonomy is one of the important components of ontology. The (semi-) automatic acquisition of taxonomy based on the actual terminology used by a community is a major step towards the creation of (full) ontology. However, several issues still need to be resolved in order to construct ontology effectively and efficiently. First of all, the knowledge acquisition bottleneck problem. Much research has been devoted to develop methods to (semi-) automatic acquisition of taxonomies from text. This type of text mining is considered by Cimiano (2006) as a process of reverse engineering because it is a task of reconstructing the world model of an author(s) from the document(s) he has written. This task is inherently complex and challenging mainly due to two reasons. First, there is only a small part of the authors' domain knowledge available (i.e., in the text) in the recreation process. Second, the 'small' domain knowledge in the text is rarely mentioned explicitly, except in a dictionary. Taxonomy learning from texts thus can be seen as working with the '*unknown symbols*' for which the appropriate sense needs to be identified as some sort of reverse engineering.

Taxonomy learning from texts is the use of content found in texts, for data mining tasks such as clustering. According to Cimiano (2006), making explicit the knowledge implicitly contained in texts is a great challenge. The (semi-)automatic taxonomy acquisition from text is basically an advanced text mining that aims to identify knowledge structure in the form of taxonomy as knowledge can be found in texts at

different levels of explicit. Thus, common methods exploits regular patterns to such as '*computer is a machine'* to discover taxonomic relations. However, it seems that the more technical and specialized the texts are, the less basic knowledge will be found stated in an explicit manner (Cimiano, 2006). Thus, conceptual clustering is an alternative technique to derive knowledge from texts. This technique is based on distributional hypothesis (Harris, 1954) that assumes that terms are similar to the extent to which they share similar linguistic contexts. In other words, Harris's hypothesis states that the meaning of words corresponds to their use in texts. Although, many techniques and methods have been implemented based on Harris' (1968) distributional hypothesis in learning taxonomy from texts, new techniques are expected to produce better quality in terms of recall and precision. Though several methods have achieved the goal in taxonomy learning from text, some draw backs still exist.

Conceptual clustering methods which have been used for taxonomy induction are hierarchical agglomerative clustering (HAC) algorithms, divisive algorithms such as Bisecting *K*-Means and Formal Concept Analysis (FCA). FCA and HAC attract a lot of attentions since the clustering result is presented in the form of a tree structure or lattice. These methods are based on Harris's distributional hypothesis (Harris, 1954) to automatically derive taxonomies from texts. Even though these methods have produced good results, there are few issues shared by these methods such as data sparseness. Cimiano (2006) states that certain contextual features which are extracted from the texts are accidental and erroneous due to missing data or grammatical mistakes, thus not corresponding to real-world or semantic similarities. On the other hand, these approaches suffer from the incapability to appropriately label the obtained clusters.

Various attempts to overcome data sparseness have been made as can be seen in Buitelaar *et al.* (2003), Cimiano *et al.*(2004a; 2004b), Reinberger and Spyns (2005) and Blohm and Cimiano (2007). Cimiano (2005) for example, introduces a novel algorithm called Guided Agglomerative Hierarchical Clustering (GAHC) and a technique called *smoothing* in another different taxonomy learning system which is based on Formal Concept Analysis (FCA). GAHC exploits WordNet and Hypernym Oracle (HO) to guide the clustering process in order to create reasonable clusters even without enough data. The HO is populated by pairs of hypernym/hyponym. In order to find the hypernym and hyponym, GAHC uses pattern matching-based approach. GAHC produces better results

compared to other unsupervised techniques such as HAC or FCA (Cimiano, 2006). The smoothing technique is used in order to overcome data sparseness by using the conditional probability to filter the extracted terms before the similarity between terms are measured using cosine metric.

However, approaches based on pattern matching as being used in GAHC suffer from a very low recall which is due to the fact that the lexico-syntactic patterns are very rare. The more technical and specialized the texts are, the less basic knowledge will be found stated in an explicit manner (Cimiano, 2006). Besides, Lian Tze and Hussein (2006) states that WordNet too often includes many rare senses while missing domain-specific senses. In this study, the WordNet has been functionally tested thoroughly and it is found that WordNet is unsuitable to be translated as representation of Malay senses. This is because translation is not merely an act of linguistic transfer, but it also involves the interaction of cultures and that transference of culture which imposes far greater problems than linguistic transfer (Elkateb *et al., 2006)*. For example, 'prophetess' exists in WordNet which is completely taboo in the Malay/Muslim world. Moreover, the smoothing technique used by Cimiano (2005) doesn't improve the results of the FCA-based approach (Ryu *et al*., 2006).

Furthermore, the discovery of taxonomy from Malay texts is conspicuous by its absence in the research literature, yet it is a very real issue. The Malay language belongs to the Austronesian family of languages (Nik Safiah, 1995). Even though the Malay language has the same Subject-Verb-Object grammatical structure as English, to Malay grammarian Malay is different from English. Azhar (1988) has shown evidence in his work from an Englishmen's opinion that Malay is different from English because Malay is a context-dependent and terse language, i.e. brief, direct to the point and 'effectively cut short' language. Ontology engineers need guidelines about the effectiveness, efficiency and trade-offs of different methods in order to decide which techniques to apply in which settings. But there is no comparative work that systematically analyze different techniques and algorithms on learning concept hierarchies from Malay text. This scenario is a major impediment for this study as it leads to *lack of guidelines* problem (Cimiano, 2006). Lack of guideline refers to the lack of research on related fields resulting in no guidelines about the effectiveness, efficiency and trade-offs of different methods to support the automatic creation of ontology from Malay text. While

this research is based on the assumption that existing taxonomy learning from texts methods can be applied to Malay texts but how should such a system be realized?

In summary, the proposed solution to these problems must be robust for the following issues that make this a taxing task as highlighted by Cimiano (2005 & 2006) as follows.

1. The texts content is noisy therefore the extracted features can be erroneous, i.e. not all derived features are correct,
2. Not all the extracted features are 'relevant' in the sense that the extracted features will help to discriminate between the different objects (terms); and
3. The assumption of completeness of information will never be fulfilled, i.e. the text collection will never be '*big enough*' to find all the possible occurrences (Ziph, 1932).

Artificial Immune System (AIS) has been the attention to many researchers in dealing with data sparseness and noise. Previous work has shown that an AIS, has attributes that fulfill these criteria. There are a number of motivations for using the immune system as inspiration for both taxonomy learning and text mining which include robustness, reliability, recognition, diversity, memory, self regulation, and learning (Dasgupta, 1999). This sophisticated natural system provides insight for solving the data sparseness and noise faced by GAHC. It is believed that these features will give the system the ability to adapt to noise and incomplete data. The fundamental challenge in learning taxonomy from Malay texts is the selection of features that are best to represent a concept. With the assumption that the existing English syntactic features and lexico-syntactic pattern may work on Malay text, the available clustering methods and the proposed approaches will be used to test this basic notion. However, it is believed that Malay language has its own unique characteristics that need to be identified to complement existing syntactic features and lexico-syntactic patterns introduced in Hearst (1992) and Cimiano (2006).

1.3 Problem Statement

The solution of the taxonomy learning from Malay texts problem can be briefly described as follows:

> Given a Malay texts, the challenges is to (semi) automatic acquisition of taxonomy from the texts while it is a norm in this approach that the extracted datasets from the text are sparse and contains noise. At the same time, the computational method must be capable of producing better taxonomy in terms of F-measure of taxonomic overlap (i.e., F_{TO}) which is defined by taxonomic recall and precision.

In light of the above statement, a robust system is required since the existing unsupervised conceptual clustering algorithms for learning taxonomy are not fault tolerance especially when handling data sparsity and noise. Therefore, it is desirable to come up with a new technique for taxonomy learning that is robust to noise and data sparsity. In order to derive knowledge in a form of taxonomy from texts (semi)-automatically, researchers have proposed conceptual clustering methods which are based on Harris (1968) distributional hypothesis. However, there are two main issues in learning taxonomy from texts.

First, a robust conceptual algorithm that can cope with data sparsity and noise and second, an efficient feature selection mechanism. The first issue is raised as the extracted attributes for some terms can be accidental or erroneous which leads to noise in data. Furthermore, data sparseness is always an issue when learning taxonomy from texts. Thus, if existing clustering technique such as bisecting K-means is used for learning taxonomy from text, only specific aspect of the whole dataset is taken into account (de Castro et al., 2007.) Artificial Immune System (AIS) has been known for its mechanism of clonal expansion and network suppression that can produce more accurate data representations in handling data sparseness. Even though AIS is known for its robustness, to date, AIS has not been applied and tested in learning taxonomy from texts.

The second issue is about answering the question of how to provide the necessary attributes (features) for every terms in an effective and efficient ways. The feature selection issue is still not solved if the source is in Malay. As potential way out of this problem, Cimiano's methodology that acquires collective knowledge from the Web

using Google can be adopted. In particular, a modified Google-based text miner can be developed to test whether it is effective in developing a better taxonomy.

A new immune-inspired algorithm may surpass the existing taxonomy learning technique and methods in precision and recall, and is able to cope with data sparsity and noise for better performance and robust taxonomy learning system. Therefore, it is necessary to demonstrate that the immune-inspired algorithm and the Google-based text miner indeed make a significant difference in building better taxonomy models. In this thesis, the question of whether the immune-inspired mechanism matters, i.e., does it lead to building better taxonomy, where by "better" we assume superior clustering and later classification performance. The following are the research questions that will be addressed to answer the above problem statement.

i. How to model AIS in learning taxonomy from texts?
ii. How to design a new hybrid method between Guided Agglomerative Hierarchical Clustering and aiNet and Clonal Selection Algorithm and Bisecting *K*-mean?
iii. How can the Artificial Immune System improve the result compared to the existing clustering methods?
iv. Is Harris's distributional hypothesis applicable in learning taxonomy from Malay texts?
v. Can hybridization of Artificial Immune System with other conceptual clustering methods leads to better result?
vi. Which affinity measurement should be used in the proposed learning system that will lead to better result?
vii. Can a Google-based Text Miner help to overcome data sparseness for better result?
viii. Can English lexico-syntactic pattern be exploited in learning taxonomy from Malay texts?
ix. How to optimize the proposed immune-inspired learning algorithm parameters?

The research questions are based on the effects of dataset (i.e., extracted from texts), metric set and feature selection and extraction technique. Based on the research questions above, the following hypotheses have been stated.

i. The proposed immune-based learning system improves the quality of generated taxonomy, compared to the available comparison methods.
ii. The hybridization of Artificial Immune System with other conceptual clustering methods leads to better result compared to existing approaches.

1.4 Aim of the Research

The aim of this research is to develop and enhance the hybrid artificial immune system that will improve the conceptual clustering performance, able to deal with data sparseness for taxonomy learning from Malay texts.

1.5 Objectives of the Study

In order to achieve the aim of this research, it is guided by the following objectives.

1. To propose a new hierarchical clustering algorithm based on the Artificial Immune System that is able to handle data sparsity robustly for learning taxonomy from Malay texts.
2. To develop an immune-inspired hybrid model for learning taxonomy from Malay texts.
3. To investigate the effect of the proposed feature selection scheme which consists of a new pseudo-syntactic pattern and Google-Based text miner for developing context distribution from Malay texts to overcome data sparsity?
4. To evaluate and validate the performance of the proposed hybrid methods and feature selection approaches with benchmark methods on three different sets of data in learning taxonomy from texts.

1.6 Scope of the Research

This research will engage in an in-depth learning taxonomy from Malay texts using bio-inspired algorithm. Thus, the research work in this study is focused to validate the effectiveness of (semi-)automatic acquisitions of taxonomy from Malay texts using immune-inspired algorithm and linguistic pattern-based approach. There are three main paradigms exploited to induce taxonomies from textual data (Buitelaar et al, 2003). The first one is the application of lexico-syntactic patterns to detect hyponymy relations. The second paradigm is based on Harris' distributional hypothesis for synonym extraction and term clustering. The third paradigm is from the information retrieval community and relies on a document-based notion of term subsumption. The scope of this study is limited to the combination of first and second paradigm. Thus, the proposed methods will be compared to the available state-of-the-art hierarchical clustering technique**.** The mathematical definitions of these measurements are presented in Chapter 4. The proposed methods will be compared to the available state-of-the-art hierarchical clustering technique**.** The mathematical definitions of these measurements are presented in Chapter 4.

This research also concerns on studying the effect of using the existing grammatical features and lexico-syntactic pattern to the performance of the learning algorithm from Malay texts. In order to extract features from the texts, two techniques will be applied in this study. The techniques are the syntactic dependency technique which has been used by Cimiano *et al.* (2006) and the lexico-syntactic pattern introduced by Hearst (1992). New grammatical features and lexico-syntactic patterns for Malay will also be investigated. In this research, only the hypernym/hyponymy or IS-A relation serves as a basis for the natural sub-concept/super-concept. Other relations than IS-A is beyond the scope of this research.

It is beyond the scope of this investigation to perform any testing regarding differing Malay Natural Language Processing tools and as such is left for future

research. As with other immune-inspired learning algorithms, the scope of discussions in this thesis is narrowed down to the working of antibodies and antigens in clonal selection and immune network.

Therefore, the scopes of the study can thus be enumerated as follows.

i. The study focuses on the (semi-) automatic acquisition of taxonomy from three Malay texts representing three different domains which are:
    a. Fiqh (Islamic Jurisprudence).
    b. Biochemistry.
    c. Information Technology (IT).
ii. This research investigates the use of Google-based feature extraction in order to overcome data sparseness.
iii. To investigate the best affinity measurement for the proposed algorithms, three affinity measurements are tested which are Hamming, Jaccard and Rand Coefficient.
iv. For the purpose of comparison, five clustering techniques are developed as follows.
    a. Guided Agglomerative Hierarchical Clustering.
    b. Bi-Secting K-Means.
    c. Hierarchical Agglomerative Clustering: Single Linkage.
    d. Hierarchical Agglomerative Clustering: Average Linkage.
    e. Hierarchical Agglomerative Clustering: Complete Linkage.
v. The quality of the produced taxonomies is measured by using measurements introduced by Cimiano (2006). The measurements are Cimiano's Taxonomic Overlap measurements which are Lexical Precision (LP), Lexical Recall (LR), Lexical F1 (F1), Precision Taxonomic Overlap Precision (P_{TO}), Recall Taxonomic Overlap (R_{TO}) and F-Measure Taxonomic Overlap (F_{TO}).

1.7 Theoretical Framework

The theoretical framework of the research is based on Maedche and colleagues' method (Kietz *et al.*, 2000). The research framework depicted in Figure 1.1 consists of four activities:

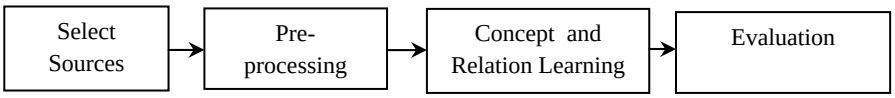

Figure 1.1: The traditional framework for learning taxonomy from texts

Activity 1. Select sources. The sources are documents which are heterogeneous in their format and contents. The documents can be domain text or generic text documents. In this study, the World Wide Web and general corpus provide the general documents since it is the largest repository in the world.

Activity 2. Preprocessing. During the pre-processing phase, the main technique used is natural language processing (NLP). Generally, the NLP involved in this study are tokenization, part-of-speech tagging, shallow syntactic analysis (or parsing), term and word extraction and stemming. Using NLP tools, meaningful features or attributes for each term are extracted. Extracting meaningful features in this study consists of preprocessing of the documents and construction of a vector space.

Activity 3. Concept and Relation Learning. Its goal is to acquire concepts which are extracted from texts by means of mainly NLP tools that use pattern-based extraction and conceptual clustering such as Bisecting *K*-Means or Hierarchical Agglomerative Clustering. In this study, the learned concepts will only be linked with *Subclass-Of* relation or also known as IS-A relation. Relations between concepts of the domain are learnt by means of pattern-based extraction.

Activity 4. Evaluation. Its goal is to evaluate the resulting taxonomy by comparing it with a reference taxonomy.

1.8 Definition of Terms

i) Robust or Robustness

The hierarchical clustering method is robust if it can produce an acceptable clustering performance regardless of data behavior such as sparsity or noise.

ii) Data Sparsity or Data Sparseness

Refers to a situation in which the extracted contextual features from texts are sparse and insufficient to identify similarities between terms. Since the contextual feature for a term is represented in binary, data sparseness also refers to a string of binaries that is populated primarily with zeros.

iii) Noise Tolerance

The ability of a method to recognize the antigens without the need of an absolute recognition as the method is tolerant to noise.

iv) Vanilla Parameter

In information technology, vanilla is an adjective meaning plain, basic or standard. In the context of this research, the vanilla parameter refers to the values which are commonly used by researchers to find optimal solution in their experiment.

1.9 Summary and Thesis Outline

This chapter has presented the motivation of the research by reviewing the background of the problem, as well as an outline of the purpose and objective of the research. In addition, the potential contribution of the research has also been highlighted. This thesis consists of eight chapters as depicted in Figure 1.2. Figure 1.2 also shows the mapping between the research objectives and the chapters in the thesis.

The structure of this thesis is as follows.

i. The first chapter gives a brief introduction to the research and briefly explains the problem statement, research objectives, research scope and brief discussion on the research methodology.
ii. Chapter 2 reveals the background to the research with a formal and mathematical definition of taxonomy. This chapter presents in detail the field of taxonomy learning from textual data, in particular describing the history and state-of-the-art in this area.
iii. Chapter 3 introduces the basics necessary to understand the immune system. The aim of this chapter is it serves to impart technical knowledge to the reader to allow for comprehension of the later chapters. Secondly it serves to provide evidence that the use of an AIS and the chosen problem domain are both justified in the context presented in this thesis.
iv. Chapter 4 describes the methodological approach of the research.
v. Chapter 5 presents the methodology, implementation and experimental results for the first taxonomy induction approach which makes use of the proposed immune-inspired algorithm named GCAINT (Guided Clustering with Artificial Immune Network).
vi. Chapter 6 presents the methodology, implementation and experimental results for the second taxonomy induction approach named CLOSAT – Clonal Selection Algorithm for Taxonomy Learning.
vii. Chapter 7 presents the methodology, implementation and experiment results for a new approach in feature selection from Malay texts by using a Google-Desktop search engine. The methodology and implementation of the automatic parameter tuning is presented in this chapter.

viii. Finally in Chapter 8, the thesis further discuss the results and interpretation of the obtained taxonomy. This chapter concludes by offering some thoughts on the use of intelligent systems for taxonomy learning in the long term.

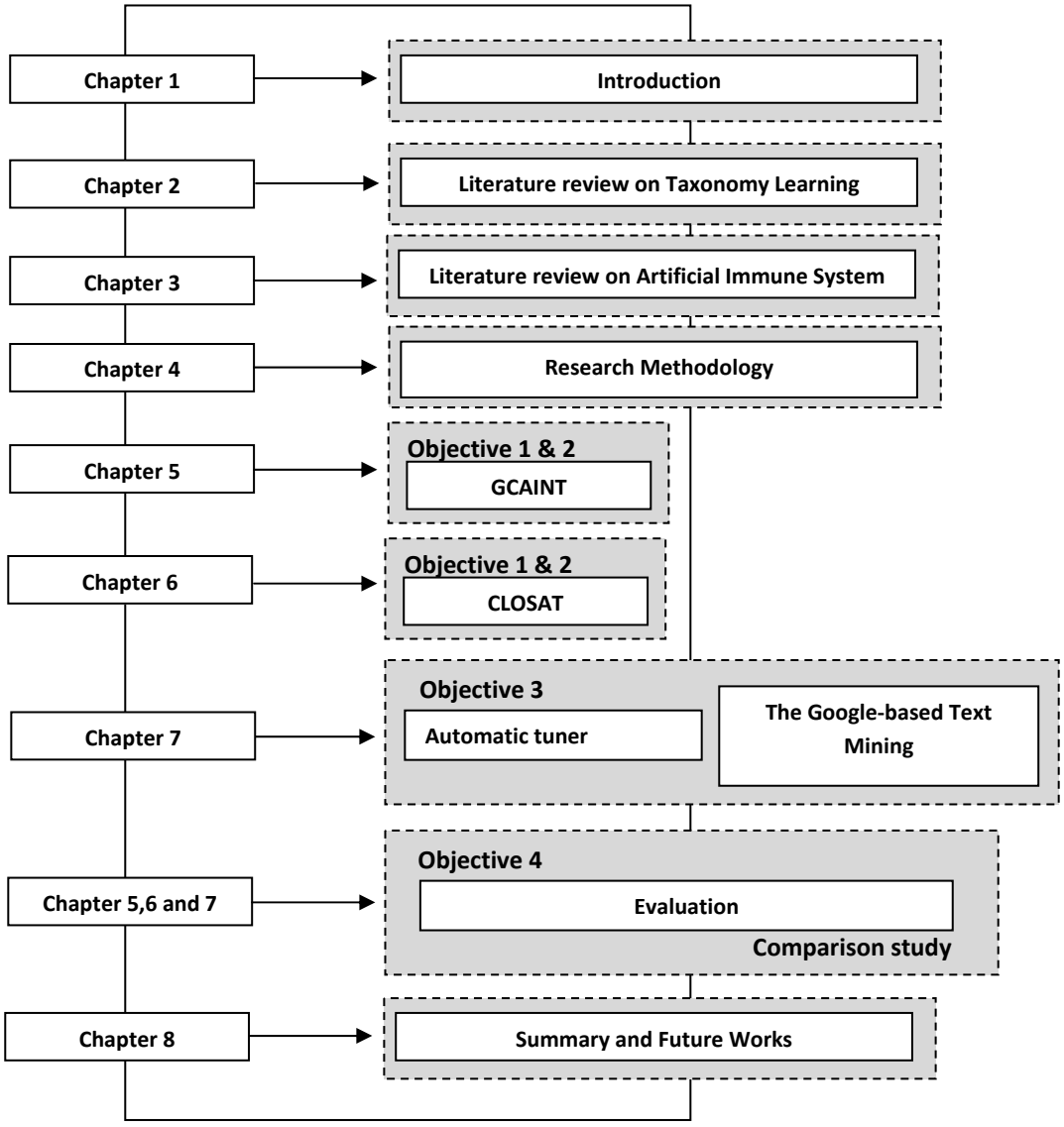

Figure 1.2: Thesis outlines and research objectives mappings

REFERENCES

(2010). Innate. EczemaNet Glossary. American Academy of Dermatology.

Acilar AM & Arslan A (2009) A collaborative filtering method based on artificial immune network Expert Systems with Applications 36 (4): 8324-8332

Aggarwal C & Yu P (2007).On Privacy-Preservation of Text and Sparse Binary Data with Sketches *Proceedings of the 2007 SIAM International Conference on Data Mining*, Minneapolis, Minnesota, 2007.

Aickelin U & Cayzer S (2002). The Danger Theory and Its Application to Artificial Immune Systems. International Conference on Artificial Immune Systems.

Al-Hammadi Y, Aickelin U & Greensmith J (2010) Performance Evaluation of DCA and SRC on a Single Bot Detection. Journal of Information Assurance and Security 5 265-275.

Alesso HP & Smith CF (2005) Developing Semantic Web Services. A K Peters.Natick, Massachusetts

Anderberg, M.R. (1973). Cluster Analysis for Applications. New York. New York: Academic Press.

Apidianaki M (2009). Data-Driven sense induction for disambiguation and lexical selection in translation. Dublin: Dublin City University.

Azhar MS (1988) Discourse-Syntax of "YANG" In Malay (Bahasa Malaysia). Dewan Bahasa dan Pustaka.Kuala Lumpur

Baedi J, Arabshahi H, Armaki MG & Hosseini E (2010). Optical Design of Multilayer Filter by using PSO Algorithm. *Research Journal of Applied Sciences, Engineering and Technology* 2 (1): 56-59.

Baroni M & Bisi. S (2004).Using cooccurrence statistics & the web to discover synonyms in a technical language. The 4th International Conference on Language Resources and Evaluation, 2004, 5, 1725-1728.

Barrière C (2004). Knowledge-Rich Contexts Discovery *Advances in Artificial Intelligence*. Belrin/Heidelberg: Springer.

Batty D, Johnson JA, Mortensen CH & Rainey KC (2010). Taxonomy. Ethnographic Thesaurus. The American Folklore Society.

Beck G & Habicht GS (1996) Immunity and the Invertebrates. Scientific American: 60-66.

Beil F, Ester M, Fraser S & Xu X (2002).Frequent term-based text clustering. *KDD '02 Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining* New York, 2002, ACM.

Bentley P (2001) *Digital Biology*. Headline. London.

Benz D. (2007). Collaborative ontology learning. Master, University of Freiburg.

Berek C & Milstein C (1988) The Dynamic Nature of the Antibody Repertoire., 1988(105):5-26. Immunology Reviews 105 5-26.

Bersini H & Varela F (1990).Hints for Adaptive Problem Solving Gleaned from Immune Networks. Proceedings of the First Conference on Parallel Problem Solving from Nature, 1990, 343-354.

Bersini H (2002). Self-Assertion versus self-recognition: a tribute to Francisco Varela.In:Timmis J & Bentley Pj, eds Proceedings of the 1st international conference on artificial immune systems (ICARIS, University of Kent at Canterbury, 2002, University of Kent.107–112.

Bezerra GB, Barra TV, Castro LNd & Zuben FJV (2005). Adaptive Radius Immune Algorithm for Data Clustering *Artificial Immune Systems*. Berlin/Heidelberg: Springer.

Bezerra GB & de Castro LN (2003).Bioinformatics Data Analysis Using an Artificial Immune Network. Proc. of Second International Conference on Artificial Immune Systems (ICARIS 2003), Edinburgh, UK, 2003, Lecture Notes in Computer Sciences, Springer, 22-33.

Bezerra GB, de Castro LN & von Zuben FJ (2004). A Hierarchical Immune Network Applied to Gene Expression Data. *Artificial Immune Systems*. Berlin / Heidelberg: Springer

Bhríde FMG, McGinnity TM & McDaid LJ (2005). Landscape Classification and Problem Specific Reasoning for Genetic Algorithms. *International Journal of Systems and Cybernetics.* 34 (9/10): 1469-1495.

Biemann C (2005) Ontology Learning from Text - a Survey of Methods. LDVForum 20 75-93.

Birattari M, Sttzle T, Paquete L & Varrentrapp K (2002). A racing algorithm for configuring metaheuristics. . Proceedings of the Genetic and Evolutionary Computation Conference. San Francisco: Morgan Kaufmann Publishers Inc.

Bisson G, Nedellec C & Canamero L (2000).Designing clustering methods for ontology building :The Mo'K workbench. Proceedings of the ECAI Ontology Learning Workshop., Berlin, 2000, 13-19.

Blohm S & Cimiano P (2007). Using the Web to Reduce Data Sparseness in Pattern-Based Information Extraction *Knowledge Discovery in Databases: PKDD 2007*. Berlin/Heidelberg: Springer.

Boicu M, Tecuci G, Stanescu B, Marcu D & Cascaval C (2001).Automatic Knowledge Acquisition from Subject Matter Experts. In Proceedings of the 2001 International Conference on Tools With Artificial Intelligence, Dallas, USA., 2001.

Borsje J. (2007). Rule Based Semi-Automatic Ontology Learning. Master, Erasmus University Rotterdam.

Box GE, Hunter JS & Hunter WG (2005) Statistics for Experimenters: Design, Innovation and Discovery. John Wiley & Sons.New York

Brewster C, Iria J, Zhang Z, Ciravegna F, Guthrie L & Wilks Y (2007).Dynamic Iterative Ontology Learning. Recent Advances in Natural Language Processing (RANLP 07), Borovets, Bulgaria, 2007.

Brewster C, Jupp S, Luciano J, Shotton D, Stevens RD & Zhang Z (2008).Issues in learning an ontology from text. Bio-Ontologies 2008: Knowledge in Biology, Toronto, ON, Canada, 2008, BioMed Central.

Brocke S, Veromaa T, Weissman I, Gijbels K & Steinman L (1994). Infections and multiple sclerosis: A possible role for superantigens? *Trends Microbiol.* 2 250-254.

Brown P, Pietra V, Pietra V & Mercer R (1991).Word-sense disambiguation using statistical methods. In . [1, 2]. 29th Annual Meeting of the Association for Computational Linguistics, Berkeley, California, USA, 1991, ACL.

Brunzel M (2007). Learning of Semantic Sibling Group Hierarchies - K-Means vs. Bi-secting-K-Means *Data Warehousing and Knowledge Discovery*. Berlin / Heidelberg: Springer.

Budi, I. & Bressan, S. (2003). Association rules mining for name entity recognition. In: the Fourth International Conference on Web Information Systems Engineering (WISE 2003) (pp. 325-328). Rome, Italy: IEEE Xplore.

Buitelaar P, Cimiano P & Magnini B (2003). Ontology Learning from Text:An Overview. *Ontology Learning from Text: methods, evaluation and applications*. Amsterdam: IOI Press.

Burgun A & Bodenreider O (2001).Aspects of the Taxonomic Relation in the Biomedical Domain. International Conference on Formal Ontology in Information Systems, Ogunquit, Maine, USA 2001, ACM, 222-233.

Burnett FM (1959) The Clonal Selection Theory of Aquired Immunity. Cabridge University Press.

Campelo F, Guimaraes FG, Igarashi H, Ramirez JA & Noguchi S (2006) A modified immune network algorithm for multimodal electromagnetic problems. IEEE Transactions on Magnetics Volume 42 (4): 1111-1114.

Caraballo SA (1999).Automatic construction of a hypernym-labeled noun hierarchy from text. 7th Annual Meeting of the Association for Computational Linguistics, College Park, Maryland 1999, In Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics, ACM, 120-126.

Cayzer S, Smith J, Marshall JAR & Kovacs T (2005).What Have Gene Libraries Done For AIS? 4th International Conference on Artificial Immune Systems (ICARIS 2005), Banff, Canada, 2005, Lecture Notes in Computer Science, 3627, 86-99.

Cha S-H, Yoon S & Tappert CC (2005).On Binary Similarity Measures for Handwritten Character Recognition. *Eighth International Conference on Document Analysis and Recognition (ICDAR'05)*, Seoul, Korea. , 2005.

Chan K, Saltelli A & Tarantola. S (1997). Sensitivity analysis of model output: variancebased methods make the difference. . Proceedings of the 29th conference on Winter simulation (Winter Simulation Conference). New York: ACM Press.

Charles WG (2000) Contextual correlates of meaning. Applied Psycholinguistics 21 505-524.

Chau DH, Myers B & Faulring A (2008). Feldspar: A System for Finding Information by Association. CHI 2008. Florence, Italy: ACM.

Chen J & Li Q (2006). Concept Hierarchy Construction by Combining Spectral Clustering and Subsumption Estimation. *Web Information Systems – WISE 2006*. Berlin / Heidelberg: Springer

Cheng Y & Church GM (2000).Biclustering of expression data. Proceedings of the 8th international conference on intelligent systems for molecular biology, 2000, 93-103.

Chung S, Jun J & McLeod D (2006).A Web-Based Novel Term Similarity Framework for Ontology Learning. Ontologies, Databases and Applications of Semantics (ODBASE) 2006 International Conference, Montpellier, France, 2006, 4725/2006, Springer Berlin / Heidelberg, 1092-1109.

Chye TS (1990) Biokimia Tumbuhan Hijau. Dewan Bahasa dan Pustaka.Kuala Lumpur

Cimiano P (2005).Ontology Learning from Text.In:Kushmerick N, Ciravegna F, Doan A, Knoblock C & Staab S, eds Machine Learning for the Semantic Web, 2005.

Cimiano P (2006) Ontology Learning and population from Text. Springer Berlin

Cimiano P, Hotho A & Staab S (2004a).Clustering concept hierarchies from text. . The fourth international conference on Language Resources and Evaluation, Lisbon, Portugal, 2004a, In Proceedings of LREC.

Cimiano P, Hotho A & Staab. S (2004b).Comparing conceptual, divisive and agglomerative clustering for learning taxonomies from text. Proceedings of the European Conference on Artificial Intelligence (ECAI), 2004b, 435-439.

Cimiano P, Staab S & Handschuh S (2004c).Towards the Self Annotating Web. 13th International World Wide Web Conference (WWW 2004), New York, 2004c, ACM.

Cimiano P, Hotho A & Staab S (2005a) Learning Concept Hierarchies from Text Corpora Using Formal Concept Analysis. Journal of Artificial Intelligence Research 24 305–339.

Cimiano P, Pivk A, Schmidt-Thieme L & Staab S (2005b). Learning Taxonomic Relations from Heterogeneous Sources of Evidence. *Ontology Learning from Text: Methods, Evaluation and Applications*. IOS Press.

Cimiano P & Staab S (2005).Learning Concept Hierarchies from Text with a Guided Agglomerative Clustering Algorithm. International Conference on Machine Learning 2005 (ICML 2005) Bonn Germany, 2005, Workshop on Learning and Extending Ontologies with Machine Learning Methods.

Cimiano P, Valker J & Studer R (2006) Ontologies on Demand? - A Description of the State-of-the-Art, Applications, Challenges and Trends for Ontology Learning from Text. Information, Wissenschaft und Praxis 57 (6-7): 315-320.

Cimiano P, Mädche A, Staab S & Völker J (2009). *Ontology Learning*. *Handbook of Ontologies*.Springer Verlag.

Coelho GP, de França FO & von Zuben FJ (2008). A multi-objective multipopulation approach for biclustering. *Artificial Immune System*. Berlin/Heidelberg: Springer.

Corcho O & Gomez-Perez A (2000). A Roadmap to Ontology Specification Languages. *Knowledge Engineering and Knowledge Management Methods, Models, and Tools*. Berlin: Springer.

D. Dasgupta ZJaFG (2003).Artificial Immune System (AIS) Research in the Last Five Years. Proceedings of the IEEE Congress on Evolutionary Computation (CEC), Canberra, Australia, 2003.

Dahab MY, Hassan HA & Rafea A (2006).TextOntoEx: Automatic Ontology Construction from Natural English Text. AIML 06 International Conference, Sharm El Sheikh, Egypt, 2006, 51-57.

Dasgupta D (1999) Artificial Immune Systems and Their Applications. Springer-Verlag.Berlin

Davies MN, Secker A, Freitas AA, Clark E, Timmis J & Flower DR (2008) Tuning amino acid groupings for GPCR classification , September 15, 2008; 24(18). Bioinformatics 24 (18): 3113-3118.

de Castro LN & von Zuben FJ (2000a). An evolutionary immune network for data clustering. 6th Brazilian Symposium on Neural Networks (SBRN 2000), Rio de Janiero, 2000a, IEEE Computer Society, 84-89.

de Castro LN & von Zuben FJ (2000b).The clonal selection algorithm with engineering applications. Proceedings of Genetic and Evolutionary Computation, Las Vegas, USA, 2000b, 36–37.

de Castro LN & von Zuben FJ (2001). aiNet: an artificial immune network for data analysis. In: Abbass Ha, Saker Ra & Newton Cs (eds.) *Data mining: a heuristic approach*. USA: Idea Group Publishing

de Castro LN & von Zuben FJ (2002) Learning and Optimization Using the Clonal Selection Principle. IEEE Transactions on Evolutionary Computation 6 (3).

de Castro LN & Timmis J (2002a).An artificial immune network for multimodal function optimization. Proceedings of the 2002 Congress on Evolutionary Computation, 2002. CEC '02., 2002a, 1, IEEE, 699 - 704

de Castro LN & Timmis J (2002b) Artificial Immune Systems: A new Computational Intelligence Approach. Springer. London

de Castro LN & Timmis J (2002c). Hierarchy and Convergence of Immune Networks: Basic Ideas and Preliminary Results. Proc. First International Conference on Artificial Immune System (ICARIS 2002).

de Castro PAD, de Franca FO, Hamilton M. F, Coelho GP & von Zuben FJ (2009) Query expansion using an immune-inspired biclustering algorithm. Natural Computing.

de Castro PAD, França FOd, Ferreira HM & Zuben FJV (2007). Applying Biclustering to Text Mining: An Immune-Inspired Approach. *Artificial Immune Systems* Berlin/Heidelberg: Springer.

de Vel O, Anderson A, Corney M & Mohay G (2001) Mining e-mail content for author identification forensics. ACM SIGMOD Record 30 (4): 55-64.

Dittami S (2009). Shapiro-Wilk Normality Test.

Dolan W, vanderwende L & Riichardson S (1993).Automatically deriving structured knowledge bases from online dictionaries. Proceedings of the Pasific Asociation for Computational Linguistics (PACLING), 1993, 5-14.

Dorigo M, Maniezzo V & Colorni A (1996). The Ant system: Optimization by a colony of cooperationg agents. . IEEE Trans Syst Man Cybernetic 26 (1): 29-41.

Drumond L & Girardi R (2008). A Survey of Ontology Learning Procedures. In: Freitas Flgd, Stuckenschmidt H, Pinto Hs, Malucelli A & Corcho Ó (eds.) The 3rd Workshop on Ontologies and their Applications . . Salvador, Bahia, Brazil, October 26, 2008: CEUR-WS.org.

Dunham MH (2003) *Data Mining: Introductory and Advanced Topics*. Pearson Education.New Jersey

Eberhart R & Kennedy J (1995). A new optimizer using particle swarm theory. *The IEEE Sixth International Symposium on Micro Machine and Human Science*, 1995.39-43.

Eberhart RC & Shi Y (1998).Comparison between genetic algorithms and particle swarm optimization.In:Porto Vw, Saravanan N, Waagen D & Eiben Ae, eds *The 7th Ann. Conf. on Evolutionary Programming*, San Diego, CA, 1998, Springer-Verlag.

Eberhart RC & Shi Y (2000).Comparing Inertia Weights and Constriction Factors in Particle Swarm Optimization. . *IEEE congress evolutionary computation.*, 2000, San Diego, CA.84-88.

Egozi O, Gabrilovich E & Markovitch. S (2008).Concept-Based Feature Generation and Selection for Information Retrieval. . The Twenty-Third AAAI Conference on Artificial Intelligence, Chicago, IL., 2008.

Eiben AE & Jelasity. M (2002). A critical note on experimental research methodology in ec. In Proceedings of the 2002 Congress on Evolutionary Computation (CEC '02),. USA: IEEE Press.

Elkateb, S., Black, W., Rodríguez, H., Alkhalifa, M., Vossen, P., Pease, A. & Fellbaum, C. (2006). Building a WordNet for Arabic In: the Fifth International Conference on Langauge Resources and Evaluation (LREC2006). Genoa, Italy.

Ercan MF (2008). A performance comparison of PSO and GA in scheduling hybrid flow-shops with multiprocessor tasks. Proceedings of the 2008 ACM symposium on Applied computing. Fortaleza, Ceara, Brazil: ACM.

Erinjeri JP, Picus D, Prior FW, Rubin DA & Koppel P (2008) Development of a Google-Based Search Engine for Data Mining Radiology Reports Journal of Digital Imaging 22 (4): 348-356.

Etzioni O, Cafarella M, Downey D, Kok S, Popescu A-M, Shaked T, Soderland S, Weld DS & Yates A (2004).Web-scale information extraction in KnowItAll. The 13th World Wide Web Conference, 2004, 100-109.

Evans R (2003).A framework for named entity recognition in the open domain. Th Recent Advances in Natural Language Processing (RANLP-2003), 2003, 137-144.

Fanelli RL (2008). A Hybrid Model for Immune Inspired Network Intrusion Detection *Artificial I mmune System*. Berlin/Heidelberg: Springer.

Fang L & Le-Ping L (2005). Unsupervised Anomaly Detection Based n an Evolutionary Artificial Immune Network *Applications on Evolutionary Computing*. Berlin / Heidelberg: Springer.

Farmer J, Packard N & Perelson A (1986) The immune system, adaptation, and machine learning. Physica 22 187-204.

Faure D & Nedellec C. (1998). A corpus-based conceptual clustering method for verb frames and ontology.

Faure D & Poibeau T (2000).First experiments of using semantic knowledge learned by ASIUM for information extraction task using INTEX.In:Staab S, Maedche A, Nedellec C & Wiemer-

Hastings P, eds 14th European Conference on Artificial Intelligence ECAI'00, Berlin, Germany, 2000, Proceedings of the Workshop on Ontology Learning.

Firth JR (1957) A synopsis of linguistic theory 1930-1955. Longman.London

Forrest S, Hofmeyr SA & Somayaji A (1997) Computer Immunology. Communications of the ACM 40 (10): 88-96.

Forrest S, Perelson AS, Allen L & Cherukuri R (1994).Self-Nonself Discrimination in a Computer IEEE Symposium on Research in Security and Privacy, Los Alamitos, USA, 1994, IEEE Computer Society Press, 202-212.

Francois O & Lavergne. C (2001) Design of evolutionary algorithms - a statistical perspective. EEE Transactions on Evolutionary Computation 5 (2): 129-148.

Ganter W & Wille R (1999) Formal Concept Analysis:Mathematical Foundations. Springer

Gao XZ, Ovaska SJ, Wang X & Chow M-Y (2009) Clonal optimization-based negative selection algorithm with applications in motor fault detection Neural Computing & Applications 18 (7): 719-729.

Geffet M & Dagan I (June 2005).The Distributional Inclusion Hypotheses and Lexical Entailment. Proceedings of the 43rd Annual Meeting of the ACL, Ann Arbor, June 2005, Association for Computational Linguistics, 107-114.

Ghani RA, Husin NM & Yim CL. (2004). Pangkalan Data Korpus DBP:Perancangan, Pembinaan dan Pemanfaatan. Available: httpGoldsby RA, Kindt TJ, Osborne BA & Kuby J (2003) *Immunology*.(5th Edition) W H Freeman.New York

Gomez-Perez A, Corcho-Garcia O & Fernandez-Lopez M (2005) Ontological Engineering. Springer-Verlag.New York

Gomez-Perez A & Manzano-Macho D (2004) An overview of methods and tools for ontology learning from texts. The Knowledge Engineering Review 19 (3): 187-212.

González, F., Dasgupta, D. & Gómez, J. (2003). The Effect of Binary Matching Rules in Negative Selection In: Genetic and Evolutionary Computation — GECCO 2003 (Vol. 2723/2003). Berlin/Heidelberg: Springer.

Goodman DE, Boggess LC & Watkins AB (2002).Artificial Immune System Classification of Multiple-class Problems. In Proc. of Intelligent Engineering Systems, 2002, ASME, 179-184.

Granitzer M, Augustin A, Kienreich W & Sabol V (2009).Taxonomy Extraction from German Encyclopedic Texts. *In Proceedings of the Malaysian Joint Conference on Artificial Intelligence*, Kuala Lumpur, Malaysia, 2009.

Greensmith J & Cayzer S (2004). An Artificial Immune System Approach to Semantic Document Classification *Artificial Immune Systems*. Heidelberg: Spinger.

Grefenstette G (1994) Explorations in Authomatic Thesaurus Construction. Kluwer

Grosjean J, Plaisant C & Bederson B (2002). SpaceTree. 1.6 ed. Maryland: Human-Computer Interaction Lab, University of Maryland.

Gruber TR (1993) A translation approach to portable ontologies. Knowledge Acquisition 5 (2): 199-220.

Guarino N & Giaretta P (1995).Ontologies and Knowledge Bases: Towards a Terminological Clarification.In:Mars Nji, ed. International Conference on Building and Sharing Very Large Scale Knowledge Bases, Amsterdam, 1995, Towards Very Large Knowledge Bases, Knowledge Building & Knowledge Sharing, IOS Press, 25-32.

Gulla JA & Brasethvik T (2008). A Hybrid Approach to Ontology Relationship Learning. *Natural Language and Information Systems*. Berlin/Heidelberg: Springer.

Hamdan AR, Nasrudin MF, Sarim HM, Zakaria MS, Yahya Y, Murah MZ & Nazri MZA (2004) Teknologi Maklumat dan Komunikasi. McGraw-Hill.Kuala Lumpur

Hamzah MP. (2006). Frasa dan Hubungan Semantik Dalam Perwakilan Pengetahuan:kesan Terhadap Keberkesanan Capaian Dokumen Melayu. Phd, Universiti Kebangsaan Malaysia.

Hamzah MP & Sembok TMT (2005). Enhancing Retrieval Effectiveness of Malay Documents by Exploiting Implicit Semantic Relationship between Words. *World Academy of Science, Engineering and Technology* 10.

Hand D, Mannila H & Smyth P (2001) Principles of Data Mining. The MIT Press

Hang X & Dai H (2004). An Immune Network Approach for Web Document Clustering. Proceedings of the 2004 IEEE/WIC/ACM International Conference on Web Intelligence. IEEE Computer Society.

Harmer PK, Williams PD, Gunsch GH & Lamont GB (2002) An artificial immune system architecture for computer security applications. IEEE Transactions on Evolutionary Computation 6 252-280.

Harris Z (1954) Distributional structure. Word 10 ((23):): 146-162.

Harris Z (1968) Mathematical Structure of Language. Wiley

Hart E, Ross P & Nelson J (1998) Producing Robust Schedules via an Artificial Immune System. Evolutionary Computation 6 (1): 61-81.

Hearst M (1992).Automatic Acquisition of Hyponyms from large Text Corpora. Proc. of 14th COLING, Nantes, France, 1992.

Hindle D (1990).Noun classification from predicateargument structures. . Annual Meeting of the Association for Computational Linguistics., 1990, In Proceedings of the Annual Meeting of the Association for Computational Linguistics.

Hoffmann GW (1986) A Neural Network Model Based on the Analogy with the Immune System. Journal of Theoretical Biology 122 33-67.

Hofmeyr S & Forrest S (2000) Architecture of an Artificial Immune System. Evolutionary Computing 8 (4): 443-473.

Holland J (1992). Genetic algorithms. Scientific American 66-72.

Hu X. (2006). Particle Swarm Optimization [Online]. Available: http://www.swarmintelligence.org [Accessed 5 Jan 2010 2010].

Hu Y, Zheng Q, Bai H, Sun X & Dang H (2005). Taxonomy Building and Machine Learning Based Automatic Classification for Knowledge-Oriented Chinese Questions. In: Huang Ds, Zhang X-P & Huang G-B (eds.) *Advances in Intelligent Computing*. Berlin/Heidelberg: Springer.

Hunt J, Timmis J, Cooke D, Neal M & King C (1999) The Development of an Artificial Immune System for Real World Applications. Artificial Immune System and Their Applications.

Hunt JE & Cooke DE (1996) Learning using an artificial immune system. Journal of Network and Computer Applications 19 189-212.

Hsu C-C, Chen C-L & Su Y-W (2007). Hierarchical clustering of mixed data based on distance hierarchy. *Information Sciences* 177 (20): 4474-4492.

Imsombut A & Kawtrakul A (2007) Automatic building of an ontology on the basis of text corpora in Thai Language Resources and Evaluation 42 (2): 137-149.

Iwanska LM, Mata N & Kruger. K (2000). Fully automatic acquisition of taxonomic knowledge from large corpora of texts. In: Shapiro Lmiasc (ed.) *Natural Language Processing and Knowledge Processing*. MIT/AAAI Press.

Jaccard, P. (1901). Étude comparative de la distribution florale dans une portion des Alpes et des Jura. . Bull Soc Vaudoise Sci. Nat., 37, 547-579.

Jain AK (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters* 31 651-666.

Janeway CA, Travers P, Walport M & Shlomchik MJ (2005) Immunobiology: the immune system in health and disease. Garland Science Publishing.

Jerne NK (1974). Towards a network theory of the immune system. Annals of Immunology (Inst Past) 125 (C): 373-389.

Jiang JJ & Conrath DW (1997).Semantic Similarity Based on Corpus Statistics and Lexical Taxonomy. International Conference Research on Computational Linguistics ROCLING X, Taipei, Taiwan, 1997.

Jiang X & Tan A (2010) CRCTOL: A semantic-based domain ontology learning system. Journal of the American Society for Information Science and Technology 61 (1).

Kavalec, M. and Svatek, V. (2005). A study on automated relation labeling in ontology learning. In Buitelaar, P., Cimiano, P., and Magnini, B., editors, *Ontology Learning from Text: Methods, Evaluation and Applications,* number 23 in Frontiers in Artificial Intelligence and Applications, pages 44-58. IOS Press.

Kelsey J & Timmis J (2003).Immune Inspired Somatic Contiguous Hypermutation for Function Optimisation.In:Cantu-Paz E, ed. GECCO 2003, Chicago, IL, USA, July 2003 2003, Lecture Notes in Computer Science, Vol 2723, Springer, pp 207-218.

Kietz J, Maedche A & Volz R (2000).A Method for Semi-Automatic Ontology Acquisition from a Corporate Intranet.In:Aussenac-Gilles N, Biébow B & Szulman S, eds EKAW'00 Workshop on Ontologies and Texts, Juan-Les-Pins, France, 2000, CEUR Workshop Proceedings.

Kim J & Bentley P (2002).Towards an Artificial Immune Systems for Network Intrusion Detection: An Investigation of Dynamic Clonal Selection. Congress on Evolutionary Computation (CEC-2002), Honolulu, 2002, IEEE, 1015-1020.

Knight T & Timmis J (2002).Multi-Layered Immune Inspired Approach to Data Mining. 4th International Conference on Recent Advances in Soft Computing, Nottingham, UK, 2002, 266-271.

Knight T & Timmis J (2003). A Multi-layered Immune Inspired Machine Learning Algorithm. In: Lotfi A & Garibaldi M (eds.) *Applications and Science in Soft Computing* Berlin/Heidelberg: Springer-Verlag.

Kuo H-C, Tsai T-H & Hung J-P (2006).Building a Concept Hierarchy by Hierarchical Clustering with Join/Merge Decision International Conference on Computational Intelligence in Economics and Finance, 2006.

Landauer T & Dumais S (1997) A solution to plato's problem: The latent semantic analysis theory of acquisition. Psychological Review 104 (2): 211-240.

Lassila O & McGuinness DL. (2001). The Role of Frame-Based Representation on the Semantic Web. *Knowledge Systems Laboratory Report* [Online]. Available: http://www.ksl.stanford.edu/KSL_Abstracts/KSL-01-02.html.

Lau H, Timmis J & Bate I (2009a).Anomaly detection inspired by immune network theory: A proposal. IEEE Congres on Evolutionary Computation (CEC09), 2009a, IEEE, 3045-3051.

Lau RYK, Lai CCL, Ma J & Li Y (2009b). Automatic Domain Ontology Extraction for Context-Sensitive Opinion Mining. 13th International Conference on Information Systems (ICIS 2009). Phoenix: Association for Information Systems.

Lhotská L, Macaš M & Burša M (2006). *PSO and ACO in Optimization Problems*. In: Al Ece (ed.) *Intelligent Data Engineering and Automated Learning (IDEAL 2006)*.(1390 – 1398). Heidelberg: Springer.

Li G, Zhuang J, Hou H & Yu D (2009).An Improved Clonal Selection Classifier Incorporating Fuzzy Clustering. International Conference on Measuring Technology and Mechatronics Automation, 2009. ICMTMA '09. , Zhangjiajie, Hunan 2009, IEEE, 179 - 182

Li Z & Tan H-Z (2006). A Combinational Clustering Method Based on Artificial Immune System and Support Vector Machine. In: Gabrys B, Howlett Rj & Jain Lc (eds.) *Knowledge-Based Intelligent Information and Engineering Systems*. Berlin Heidelberg: Springer-Verlag.

Lin D & Pantel P (2002).Concept Discovery from Text. Conference on Computational Linguistics (COLING-02), Taipei, Taiwan, 2002, In Proceedings of Conference on Computational Linguistics (577-583.

Liu R, Sheng Z & Jiao L (2009).Gene transposon based clonal selection algorithm for clustering. Proceedings of the 11th Annual conference on Genetic and evolutionary computation, Montreal, Québec, Canada 2009, ACM, 1251-1258

Looks M, Levine A, Covington GA, Loui RP, Lockwood JW & Cho YH (2007).Streaming Hierarchical Clustering for Concept Mining. *Aerospace Conference, 2007 IEEE* Big Sky, MT 2007, IEEE.

Maedche A, Pekar V & Staab S (2002). Ontology learning part One: On Discovering Taxonomic Relations from the Web. *Web Intelligence*. Berlin: Springer.

Maedche A & Staab A (2001a) Ontology learning for the semantic web. IEEE Intelligent Systems 16.

Maedche A & Staab S (2000). Discovering Conceptual Relations from Text.In:W.Horn, ed. ECAI 2000. , Berlin, 2000, Proceedings of the 14th European Conference on Artificial Intelligence, IOS Press, 21-25.

Maedche A & Staab S (2001b).Ontology Learning for the Semantic Web. Special Issue on the Semantic Web. IEEE Intelligent Systems, 2001b, 16.

Madche, A. (2002). Ontology Learning for the Semantic Web. Kluwer Academic Publishers.

Madche, A. & S., S. (2002). Measuring similarity between ontologies. In: Euroepean Conference on Knowledge Acquisition and Management (EKAW) (pp. 251-263).

Maio CD, Fenza G, Loia V & Senatore S (2009).Towards an automatic fuzzy ontology generatio. Proceedings of the 18th international conference on Fuzzy Systems, Jeju Island, Korea 2009, 1044-1049.

Makrehchi M & Kamel MS (2007). Automatic Taxonomy Extraction Using Google and Term Dependency. 2007 IEEE/WIC/ACM International Conference on Web Intelligence. Silicon Valley, USA: IEEE.

Makrehchi M & Kamel MS (2007).Automatic taxonomy extraction using google and term dependency. the IEEE/WIC/ACM International Conference on Web Intelligence (WI '07), Washington, DC, USA., 2007, IEEE Computer Society, 321–325.

Mann HB & Whitney DR (1947). On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. *Annals of Mathematical Statistics* 18 (1): 50-60.

Mann PS (2007) Introductory Statistics John Wiley & Sons.New Jersey

Marrack P & Kappler J (1994). Subversion of the immune system by pathogens. Cell 76 (2): 323-332.

Matzinger P (2002) The Danger Model: A Renewed Sense of Self. Science 296 301-305.

Mayer G. (2006). Immunology - Chapter One: Innate (non-specific) Immunity. *Microbiology and Immunology On-Line Textbook* [Online]. Available: http://pathmicro.med.sc.edu/ghaffar/innate.htm.

McDonald J (2009). Handbook of Biological Statistics. 2nd ed. ed. Baltimore, Maryland: Sparky House Publishing.

Melz R, Ryu P-M & Choi K-S (2006). Compiling large language resources using lexical similarity metrics for domain taxonomy learning. In: (2006). 5th International Conference on Language Resources and Evaluation. Genoa, Italy.

Michalsky R (1980) Knowledge Acquisition through conceptual clustering: a theoretical framework and algorithm for partitioning data into conjunctive concepts. International Journal of Policy Analysis and Information Systems 4 (3): 219-243.

Miller G (1995) Wordnet: a lexical database for english. Communication of the ACM 38 (11): 39-41.

Miller G & Charles W (1991) Contextual correlates of semantic similariry. Language and Cognitive Processes 6 1-28.

Mitchell TM (1997) Machine Learning. McGraw Hill

Mueller, C. (2004). Data Clustering. In: Indiana University.

Nanas N & Roeck Ad (2009) Autopoiesis, the immune system, and adaptive information filtering. Natural Computing 8 (387-427).

Navigli R, Velardi P, Cucchiarelli A & Neri F (2004).Quantitative and qualitative evaluation of the OntoLearn ontology learning system. *Proceeding COLING '04 Proceedings of the 20th international conference on Computational Linguistics* Stroudsburg, PA, USA, 2004, Association for Computational Linguistics 1043.

Nazri MZA, Shamsudin SM, Abu Bakar A & Abdullah S (2009).A Hybrid Approach for Learning Concept Hierarchy from Malay Text Using GAHC and Immune Network In:Andrews Ps, ed. 8th International Conference on Artificial Immune Systems, York UK, 2009, Lecture Notes in Computer Science, Springer-Berlin/Heidelberg.

Neal M, Hunt J & Timmis J (1998).Augmenting an artificial immune network. In Proc. of Int. Conf. Systems and Man and Cybernetics, San Diego, California, 1998, IEEE. .3821-3826.

Neshati M, Abolhassani H & Fatemi H (2009). Automatic Extraction of IS-A Relations in Taxonomy Learning. In: Sarbazi-Azad H, Parhami B, Miremadi S-G & Hessabi S (eds.) *Advances in Computer Science and Engineering*. Berlin Heidelberg: Springer.

Nik Safiah K. (1995) Malay Grammar for Academics and Professionals. Dewan bahasa dan Pustaka.Kuala Lumpur

Novacek V. (2005). Ontology Learning. Diploma, Masarykova Univerzita V Brne.

Obitko M, Snasel V & Smid J (2004).Ontology Design with Formal Concept Analysis.In:Snasel V & Belohlavek R, eds Proceedings of the CLA 2004 International Workshop on Concept Lattices and their Applications, Ostrava, Czech Republic, 2004, CEUR Workshop Proceedings, 110, CEUR-WS.org.

Ochiai, A. (1957). Zoogeographic studies on the soleoid fishes found in Japan and its neighbouring regions. . Bull. Jpn. Soc. Sci. Fish. 22: 526-530.

Osswald R & Petersen W (2002). Induction of classifications from linguistic data. ECAI'02 Workshop. Universit´e de Lyon.

Pang W & Coghill GM (2007).Modified clonal selection algorithm for learning qualitative compartmental models of metabolic systems. Genetic And Evolutionary Computation Conference, London, United Kingdom 2007, ACM.

Pasca M (2004).Acquisiotion of categorised names entities for web search. Proceedings of the Conference on Information and Knowledge Management (CIKM), 2004, 137-145.

Pereira F, Tishby N & Lee L (1993).Distributional clustering of english words. . 31st Annual Meeting of the Association for Computational Linguistics, 1993, In Proceedings of the 31st Annual Meeting of the Association for Computational Linguistics, 183-190.

Perelson AS & Oster GF (1979). Theoretical studies of clonal selection: Minimal antibody repertoire size and reliability of self- non-self discrimination. *J. Theoret. Biol.* 81 645-670.

Petersen W. (2001). A Set-Theoretical Approach for the Induction of Inheritance Hierarchies [Online]. Elsevier Science. Available: http://www.elsevier.nl/locate/entcs/volume51.html [Accessed].

Pickthall MM (1930) *The Meaning of the Glorious Koran.* New York

Plaisant C, Grosjean J & Bederson B (2002). SpaceTree. University of Maryland.

Purves WK, Sadava D, Orians GH & Heller. HC (2006) Life: The Science of Biology. W. H. Freeman

Quan Thanh T, Siu Cheung H, A. C. M. F & Tru Hoang C (2009) Automatic Fuzzy Ontology Generation for Semantic Web. IEEE Transactions on Knowledge and Data Engineering 18 (6): 42 - 856.

Quan TT, Hui SC & Cao TH (2004).A Fuzzy FCA-based Approach to Conceptual Clustering for Automatic Generation of Concept Hierarchy on Uncertainty Data. Proc. of the 2004 Concept Lattices and and Their Applications Workshop 2004, 1-12.

Rahmani AT & Helmi BH (2008).EIN-WUM: an AIS-based algorithm for web usage mining. Proceedings of the 10th annual conference on Genetic and evolutionary computation, Atlanta, GA, USA 2008, 291-292.

Reinberger M-L & Daelemans W (2010). Is Shallow Parsing Useful for Unsupervised Learning of Semantic Clusters? . *Computational Linguistics and Intelligent Text Processing*. Berlin / Heidelberg: Springer.

Reinberger M-L & Spyns P (2005). Unsupervised Text Mining for the learning of DOGMA-inspired Ontologies. In: Buitelaar P, Cimiano P & Magnini B (eds.) *Ontology Learning from Text: Methods, Applications and Evaluation*. Amsterdam: IOS Press.

Ritter A, Soderland S & Etzioni O (2009).What Is This, Anyway: Automatic Hypernym Discovery. *Proceedings of the 2009 AAAI Spring Symposium on Learning by Reading and Learning to Read* 2009, Association for the Advancement of Artificial Intelligence.

Ronghua D & Yue C (2010). *A Promoted Global Convergence Particle Swarm Optimization Algorithm Advancing Computing, Communication, Control and Management*.(23-30). Berlin/Heidelberg: Springer.

Ruenes DS. (2007). Domain Ontology Learning from the Web. PhD, Universitat Politècnica de Catalunya.

Ryu P-M, Kim J-H, Nam Y, Huang J-X, Shin S, Lee S-M & Choi K-S (2006).Toward Domain Specific Thesaurus Construction:Divide-and-Conquer Method. International WordNet Conference(GWC-06), 2006, The Global WordNet Association, 69-84.

Sabou M (2004). From Software APIs to Web Service Ontologies: A Semi-automatic Extraction Method *The Semantic Web – ISWC 2004*. Berlin/Heidelberg: Springer.

Sabou M (2005). *Learning web service ontologies automatic extraction method and its evaluation*. In: Buitelaar P, Cimiano P & Magnini B (eds.) *Ontology learning from text: methods, applications and evaluation*.Amsterdam: IOS Press.

Salton G & McGill M (1983) Introduction to Modern Information Retrieval. McGraw Hill Book Co.New York

Sanchez D & Moeno A (2005).Web Scale taxonomy learning.In:Biemann C & Pass G, eds Proceedings of the workshop on Extending and Learning Lexical Ontologies using Machine Learning Methods, 2005.

Sanderson M & Croft B (1999).Deriving concept hierarchies from text. Proceedings of the SIGIR Conference on Research and Development in Information Retrieval, 1999, 206-213.

Santos RMZD (ed.) 1999. Immune Responses: Getting Close to Experimental Results with Cellular Automata Models: World Scientific Publishing Company.

Schmid H (2000) Lopar: Design and implementation. . Arbeitspapiere des Sonderforschungsbereiches 340 (149).

Schickel-Zuber V & Faltings B (2007).Using hierarchical clustering for learning theontologies used in recommendation systems. *The Thirteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD 2007*, San Jose, California, 2007.599-608.

Schmitz P (2006).Inducing Ontology from Flickr Tags. the 15th International World Wide Web Conference 2006, Edinburgh, UK., 2006, IW3C2.

Secker A. (2006). Artificial Immune Systems for Web Content Mining: Focusing on the Discovery of Interesting Information. PhD, University of Kent.

Secker A, Davies MN, Freitas AA, Timmis J, Clark E & Flower DR (2009) An Artificial Immune System for Clustering Amino Acids in the Context of Protein Function ClassificatioDe castron Journal of Mathematical Modelling and Algorithms 8 (2): 103-123.

Secker A, Freitas AA & Timmis J (2008) AISIID: An artificial immune system for interesting information discovery on the web. Applied Soft Computing 8 885-905.

Sevaux M & Sorenson K (2002).A genetic algorithm for Robust Schedules. Proceedings of the Eight International Workshop on Project Management and Schedulling, Valencia, 2002, pp 330-333.

Shamsfard M & Barforoush AA (2003) The state of the art in ontology learning: a framework for comparison. The Knowledge Engineering Review 18 (4): 293-316.

Shi XH, Liang YC, Lee HP, Lu C & Wang LM (2005). An improved GA and a novel PSO-GA-based hybrid algorithm. *Inf. Process. Lett.* 93 (5): 255-261.

Shoenfeld Y (2004). The idiotypic network in autoimmunity: antibodies that bind antibodies that bind antibodies. *Nature Medicine* 10 17-18.

Shokooh-Saremi M, Nourian M, Mirsalehi M & Keshmiri H (2004). Design of multilayer polarizing beam splitters using genetic algorithm. , 233: . *Optic. Commun.* 233 57-65.

Shapiro S & Wilk M (1965) Analysis of variance test for normality (complete samples). Biometrika Vol 52 pp 591-611.

Smucker MD, Allan J & Carterette B (2007). A Comparison of Statistical Significance Tests for Information Retrieval Evaluation. 16tn ACM Conference on Information and Knowledge Management. Lisbon, Portugal: ACM.

Sokal, R.R. & Michener, C.D. (1958). A statistical method for evaluating systematic relationships. 38: 1409-1438. . Univ. Kans. Sci. Bull.

Soriano JMG (2005). Re-ranking of Yahoo Snippets with the JIRS Passage Retrieval System.

Sorensen, T. (1948). A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analyses of the vegetation on Danish commons. K. Dan. Vidensk. Selsk. Biol. Skr., 5, 1-34.

Spellward P & Kovacs T (2005).On the contribution of gene libraries to artificial immune systems. Proceedings of the 2005 conference on Genetic and evolutionary computation, Washington DC, USA 2005, ACM, 313-319.

Sporleder C (2002).A galois lattice based approach to lexical inheritance hierarchy learning. 15th Eureopean Conference on Artificial Intelligence, 2002, In Proceedings of the ECAI 2002 Workshop on Machine Learning and Natural Language Processing for Ontology Engineering.

Spyns P & Reinberger M-L (2005).Lexically Evaluating Ontology Triples Generated Automaticallu From Text. Proc. of the second European Conference on the Semantic Web, Heraklion, Greece, 2005, Springer Berlin/Heidelberg.

Steinbach M, Karypis G & Kumar V (2000).A Comparison of Document Clustering Techniques. *The Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Boston, MA, USA, 2000, ACM.

Steinman MD, L., (1994), Autoimmunity and Neurological Disorders, URL: http//aspin.asu.edu/msnews/autoimm.htm.

Stoica E & Hearst MA (2004).Nearly-automated metadata hierarchy creation. Human Language Technology Conference Boston, Massachusetts, 2004, ACM.

Studer R, Benjamins R & Fensel D (1998) Knowledge Engineering: Principles and Methods. Data and knowledge engineering 25 161-197.

Surowiecki J (2004) The Wisdom of Crowds: Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business, Economies, Societies and Nations. DoubleDay Books

Swartout B, Patil R, Knight K & Russ T (1997).Toward Distributed Use of Large-Scale Ontologies, . AAAI Spring Symposium on Ontological Engineering, Stanford University, California, 1997, 38-148.

Tauber, A. I. 2002. The biological notion of self and non-self. Stanford Encyclopedia of Philosophy Online: http://plato.stanford.edu/entries/biology-self.

Tang N & Vemuri VR (2005).An artificial immune system approach to document clustering. Proceedings of the 2005 ACM symposium on Applied computing, Santa Fe, New Mexico 2005, ACM.

Thalib, M. (1997). Fiqh Nabawi. Kuala Lumpur: Darul Nu'man.

Teng G & Papavasiliou FN (2007) Immunoglobulin Somatic Hypermutation. Annu. Rev. Genet 41 107-120.

Timmis J. (2000). Artificial immune systems: A novel data analysis technique inspired by the immune network theory. PhD, University of Wales.

Timmis J & Neal M (2001) A resource limited artificial immune system for data analysis. Knowledge-Based Systems 14 (3-4): 121-130.

Timmis J, Neal M & Hunt. J (2001) An Artificial Immune System for Data Analysis. BioSystems 55 143 - 150.

Tonegawa S (1983) Somatic generation of antibody diversity. Nature 302 575-581.

Tubbs, J.D. (1989). A note on binary template matching. Pattern Recognition, 22(4), 359-365.

Turnbull B, Blundell B & Slay J (2006) Google Desktop as a Source of Digital Evidence. International Journal of Digital Evidence 5 (1).

Turney PD (2001). Mining the web for synonyms: Pmi-ir versus lsa on toefl. the 12th European Conference on Machine Learning, 2001, 491 - 502.

Twycross J & Cayzer S (2003).An Immune-Based Approach to Document Classification International Conference on Intelligent Information Processing and Web Mining System, Zakopane, Poland, 2003, 33-48.

Tze LL & Hussein N (2006).Fast Prototyping of a Malay WordNet System. Summer School Workshop, Bangkok, Thailand, 2006, Proceedings of the Language, Artificial Intelligence and Computer Science for Natural Language Processing Applications (LAICS-NLP), 13-16.

Uschold. M (1998) Knowledge level modelling: Concepts and terminology. . Knowledge Engineering Review 13 (1).

Velardi P, Cucchiarelli A & Pétit M (2007a) A Taxonomy Learning Method and its Application to Characterize a Scientific Web Community. IEEE Transactions On Knowledge and Data Engineering 19 (2): 180-191.

Velardi P, Navigli R, Cuchiarelli A & Neri F (2005).Evaluation of OntoLean, a methodology for automatic population of domain ontologies., 2005.

Velardi P, Navigli R & Petit M (2007b).Semantic Indexing of a Competence Map to support Scientific Collaboration in a Research Community. the 20th international joint conference on Artifical intelligence, Hyderabad, India 2007b.

von Behring E & Kitasato S (1890). Ueber das Zustandekommen der Diphtherie-Immunitat und der Tetanus-Immunitat bei Thieren. Deutscbe Medizhiscbe Wocbenschift 16 1113-1114.

Waltz DL (2006) AI's 10 to Watch. IEEE Intelligent Systems 21 (3): 5-14.

Watkins A & Timmis J (2004). Exploiting Parallelism Inherent in AIRS. *Proceedings of the 3rd International Conference on Artificial Immune Systems (ICARIS 2004)*. Berlin/Heidelberg: Springer.

Weissman, I.L. & Cooper, M.D. (1993). How the Immune System Develops. Scientific American 269, 33–40

Widdows D & Dorow B (2002).A graph model for unsupervised lexical acquisition. the 19th international conference on Computational linguistics Taipei, Taiwan 2002.

Wierzchon S & Kuzelewska U (2002).Stable Clusters Formation in an Artificial Immune System.In:Timmis J & Bentley Pj, eds Proceedings of the 1st International Conference on Artificial Immune Systems (ICARIS), 2002, University of Kent at Canterbury, University of Kent at Canterbury Printing Unit, 68-75.

Wolf JD (2009). What Does an Analysis of Popular Search Queries Reveal About Today's Information Needs?. Master Thesis, Bowie State University.

Wong K-F, Li W, Xu R & Zhang Z-s (2010) Introduction to Chinese natural language processing Machine Translation.

Xiaoshu H & Honghua D (2004). An Immune Network Approach for Web Document Clustering. Proceedings of the 2004 IEEE/WIC/ACM International Conference on Web Intelligence. IEEE Computer Society.

Xu L, Mo H, Wang K & Tang N (2006). Document Clustering Based on Modified Artificial Immune Network *Rough Sets and Knowledge Technology*. Berlin / Heidelberg: Springer

Yeh J-h & Sie S-h (2006). Towards Automatic Concept Hierarchy Generation for Specific Knowledge Network. In: M.Ali & R.Dapoigny (eds.) *Advances in Applied Artificial Intelligence*. Berlin/Heildelberg: Springer.

Younsi R & Wang W (2004). A New Artificial Immune System Algorithm for Clustering *Intelligent Data Engineering and Automated Learning – IDEAL 2004*. Berlin / Heidelberg: Springer.

Yu Z, Hong Z & Ling-dong K (2009).Research of Coal-Gas Outburst Forecasting Based on Artificial Immune Network Clustering Model. Second International Workshop on Knowledge Discovery and Data Mining (WKDD 2009), 2009, IEEE, 23-27.

Zavitsanos E, Paliouras G & Vouros. G. (2006). Ontology Learning and Evaluation: A survey [Online]. Software and Knowledge Engineering Laboratory (S.K.E.L.). Available: http://www.ontosum.org/static/Publications [Accessed 24 Mac 2007].

Zeidenberg M (1990) *Neural Network Models in Artificial Intelligence*. Ellis Horwood Ltd

Zhang C & Yi Z (2009) Tree structured artificial immune network with self-organizing reaction operator Neurocomputing 73 (1-3): 336-349

Zhang J & Liang Y (2009). An Anomaly Detection Immune Model Inspired by the Mechanism of DC- T Cell Interaction *Applied Computing, Computer Science, and Advanced Communication*. Berlin Heidelberg: Springer.

Zheng H, Hu X, Si X & Yang W (2008). A Novel Object Detection Approach for Satellite Imagery Based on Danger Theory. First International Conference on Intelligent Networks and Intelligent Systems, 2008 (ICINIS '08), 2008, IEEE.

Zhou GD (2003).Modeling of long distance context dependency in Chinese. The second SIGHAN workshop on Chinese language processing Sapporo, Japan, 2003, ACM, 71 - 77

Zhuang Y, Li X, Xu B & Zhou B (2009).Information Security Risk Assessment Based on Artificial Immune Danger Theory. 2009 Fourth International Multi-Conference on Computing in the Global Information Technology, Cannes/La Bocca, France 2009, IEEE.

Ziph G (1932) Selective studies and the principles of relative frequency in language. Cambridge