Summary: Literature Review about Artificial Immune System for Finding Relevant Information through Political Weblog

Ahmad Nadzri Muhammad Nasir
Intelligence Software Engineering Lab
Faculty of Computer Science and Information System
Universiti Teknologi Malaysia
anmn85@gmail.com

Ali Selamat
Intelligence Software Engineering Lab
Faculty of Computer Science and Information System
Universiti Teknologi Malaysia
aselamat@utm.my

Md Hafiz Selamat
Intelligence Software Engineering Lab
Faculty of Computer Science and Information System
Universiti Teknologi Malaysia
mhaﬁz@utm.my

ABSTRACT
Currently, there are many weblog that have been posting many story and news about political issues in Malaysia. Therefore, it draws Internet space as an area for political bloggers to be involves in cyber war. There are many factors that contribute to this cyber war such as opinions from bloggers, spreading of rumors and information through cyberspace. Most of the political leaders from political parties in Malaysia are using electronic media such as portal, weblogs, social networking sites and etc. With these types of applications, the bloggers can give their opinions, comments and information about political issues in Malaysia. This scenario helps to grow the population of political bloggers and blog readers in Malaysia. In this paper, we introduce an adaptive model of web mining using Artificial Immune System (WMAIS) to retrieve the list of URLs from website and web pages that relevant information on political issues in Malaysia through political weblog. This model also can be used to identify the social network of political weblog and how the social network can help on widen the blogger readerships.

Keywords
Artificial Immune System, Web Mining, Social Network, Political Weblog.

1. INTRODUCTION
The current state of affairs in Malaysia turned politics are one of the critical and hot subjects in Malaysia these days. This is because of the medias that always publishing some news and causing discussion about the political issues towards the Malaysians citizenry. Nowadays, media from the Internet such as alternative media and weblogs has becomes a reliable sources that delivers new information and news about political situation in Malaysia for readers [1]. This is because the mainstream media have provided slow release of official news and the readers always asking the credibility and reliability of news from the mainstream media [1, 2]. From these reasons, many political leaders and political analysts have become bloggers that can give their own opinions, views and comment about political issues in Malaysia. Weblog also can become a space for political leaders to become defensive to themselves when facing up accusation from other party. Indirectly, it will encourage the political bloggers especially political leaders to involve in cyber war. Therefore, the blogging activity has become one of current trend politic in Malaysia for political leaders.

The characteristics of weblog also one of the reasons why political leaders and political analysts go for blogging such as minimum cost, easy to maintain, customizable and offers anyone with an agenda a platform to preach [2]. Since blogging went mainstream, it makes arises of weblog service companies that provided a services and space for everyone to jump into blogging activity [2] such as Blogspot [3], Wordpress.com [4], Livejournal [5] and etc. With these kinds of services, the bloggers who are the political leaders and people that involve in politics can do blogging for spreading news, opinions and information about political issues in Malaysia.

Political weblogs in Malaysia have their own blog readers which they like to read more news and information about political situation in Malaysia. These blog readers will search more reliable and relevant information from famous weblog and political leaders’ weblog. After they have read the first weblog, they want to discover more information and news about political issues by surfing other weblog that have linkage with the first weblog. As the result, the blog readers are trying to widen their blog readership to find and discover relevant information about political issues in Malaysia.

In this paper, we want to introduce a web mining tool that using Artificial Immune System (AIS) called web mining using AIS (WMAIS). Result from this tool will help blog readers to discover and find relevant news and information about political issues in Malaysia through political weblog. This paper is organized according to 4 sections. Section 2 explains on the motivation of this paper. Section 3 explains on literature review that gives description on other components that related with WMAIS. Section 4 we have come out with the brief process of WMAIS and the outcome of WMAIS’s result. In Section 5, we conclude all the ideas for conclusion and discussion of this paper.

2. MOTIVATION
Before doing this research, it is worth motivating this research by believing that our reasons for doing this research is to help blog readers to find relevant information and widen blog readership about political issues in Malaysia through political weblog. The motivation of developing WMAIS is to improve other searching technique called AISIID (An artificial immune system for interesting information discovery on the web) [6]
that solve the problem that have been faced by popular search engine like Yahoo [7], Google [8] and etc. In [6], the authors said that current search techniques contain an inherent problem such as the result of current search engine provide information that user already knows and not provide information that is unexpected by users. This AISIID is using AIS technique to evaluate relevant information on the Internet and text surrounding technique for crawling for the relevant URL in the website.

As we have mentioned before, our motivation to develop WMAIS is when we realized that there are disadvantages in AISIID. We need to improve the technique in AISIID because we believed the result from AISIID is not really relevant to find information about political issues in Malaysia. After doing some research on AISIID, we have met decision that our WMAIS need to change the process in AISIID and to improve the crawling technique for finding relevant information and URL. We still implementing AIS process in WMAIS because of its abilities that we discussed in Section 3. The abilities that embedded into WMAIS such as learning capability for learning the user needs to get information what the user wants and evolving process to find more relevant information about political issues in Malaysia from the Internet.

3. LITERATURE REVIEW
3.1 Weblog
Weblog or blog is one of the web applications used for compromising blog posts or content written by the blogger where the content of the blog can be categorized and sorted in reverse chronological order [9]. Weblog is using Web 2.0 technology and this technology make the weblog become more interactive. This is because of the abilities make Internet users much easier to communicate, sharing information, cooperate with other users and increase creativity [10]. The function in weblog can be as a personal diaries, technical advice columns, sports commentary, celebrity gossip, political information and etc. These functions make weblog to be categorized into several types for examples personal blog, corporate blog, political blog, video blog and etc.

Figure 1. Examples of political weblog.

There are many web service companies that give free account for Internet users to active in blogging with laws that need to be obeyed and easier instructions to create a blog in a few minutes. The services that have been provided make the weblog are easier to be maintained and customizable. Figure 1 shows examples of political weblog from political leaders and independent bloggers.

3.2 Web Mining
The widely usage of Internet make the increasing of Internet users. According to the Internet World Stats, there are about 14,904,000 Internet users in Malaysia in 2008 [11]. It also makes the overload of information sources on the World Wide Web. It has become increasingly necessary for users to utilize automated tools in find the desired information resources, track and analyze their usage patterns. These factors give rise to the necessity of creating server side and client side intelligent systems that can effectively mine for knowledge.

The web mining is an application that can discover knowledge from huge amounts of web pages, has become an urgent research area in computer science [12]. There is also a work from [13] said that web mining is use of data mining techniques to automatically discover and extract information from World Wide Web documents and services. There are three types of web mining:

- Web content mining used to recognize pattern of text and multimedia such as picture, video and etc.
- Web structure mining used to capture the pattern for hyperlink structure on the Internet.
- Web usage mining used to identify log records of website.

In this paper, we focused more on web content mining because its ability that can recognize pattern of text in each web page on the Internet. Web content mining is usually being used to mine the content of web pages and improve the content search of other tools like search engine. Web content mining can be described as the automatic search of information resource available online, and involves mining web data contents [14]. There are technologies that are usually be used in web content mining such as natural language processing and information retrieval [15]. In web content mining, there are two types of approaches i.e. agent based approach and database approach. Each approach has techniques that can be suited in web content mining. Figure 2 shows the types of approach and techniques in web content mining.

In our case, we have applying concept of web content mining in our WMAIS to extract and recognize pattern of information from the weblog. The WMAIS also use agent based approach with intelligent search agent to learn the user needs and discover the relevant information about political issues in Malaysia. The reason we choose intelligent search agent is because it used agents for searching relevant information and
these agents help to organize and interpret the discovered information from the Internet.

3.3 Artificial Immune System

From the previous sub section, we have decided to use intelligent agents search to search and discover relevant information about political issue in Malaysia through political weblog. After we have studied several algorithms, we believed that the artificial immune system (AIS) is a suitable algorithm that has suited into our WMAIS. This is because of the abilities and characteristics of AIS to solve the problems in the current search engines. AIS is an adaptive system that have been inspired by theoretical immunology and observed immune functions, principles and models, which are applied to problem solving [16]. The design of AIS is an inspiration was taken from the two most influential biologically motivated by computing paradigms which are artificial neural networks (ANN) and evolutionary algorithm (EA) [16]. In this chapter, we also discuss the idea of creation AIS algorithm and it process flow.

Initialization: create an initial random population of individuals (P).

Antigenic presentation: for each antigenic pattern, do:

Affinity evaluation: present it to the population (P) and determine its affinity with each element of the population (P);

Clonal selection and expansion: select \( n_1 \) highest affinity elements of (P) and generate clones of these individuals proportionally to their affinity with the antigen: the higher the affinity, the higher the number of copies and vice versa;

Affinity maturation: mutate all these clones with a rate inversely proportional to their affinity with the input pattern: the higher the affinity, the smaller the mutation rate, and vice versa. Add these mutated individual to the population (P) and re-choose the best individual to be kept as the memory \( m \) of the antigen presented;

Metadynamics: replace a number \( n_x \) of individuals with low affinity by (randomly generated) new ones;

Cycle: repeat step 2 until a certain stopping criteria is met.

Figure 3. Clonal selection algorithm.

The basic algorithm of AIS is based on immune system in human body. Immune system in human body is a system that consists of a complex set of cells and molecules. This immune system has natural, rapid and effective defense mechanism to helps human body from being attacked by foreign invaders such as viruses, pathogen and etc [6]. The human immune system is divide by two line defense i.e. innate immune system and adaptive immune system. In this paper, process in the WMAIS is applying on the process flow of adaptive immune system. As we know our bodies always being attacked by antigens that can stimulate the adaptive immune system. Adaptive immune system is a system that produce antibody in response to determine infectious agent [16]. The presence of antibodies in an individual can reflects to infection which that individual has already been exposed. In adaptive immune system, there are cells called B-cells that active in this system and capable to develop an immune memory which is able recognizing the same antigenic stimulus when it presented to the organism again. This avoids the return of the disease within the organism and can improve itself with each encounter of a given antigen. The characteristics and abilities of human immune system has also influence the AIS because of its performance i.e. pattern recognition, self organization, diversity, autonomy, distributivity, dynamically changing coverage, noise tolerance and etc.

There are several algorithms in AIS that can be used to solve problems such as positive selection, negative selection, clonal selection, danger theory and etc. The clonal selection algorithm is an algorithm used process on how the adaptive immune system copes with pathogenic micro-organisms. This algorithm is using B-cells to model the algorithm that involves adaptability via mutation. Briefly, when a B-cell receptor recognizes a non-self antigen with a certain affinity, it will be selected to proliferate and produces antibodies in high volumes. Clonal selection process with the affinity proportional mutation can promotes to greedy search where it can be used in pattern recognition and optimization [18, 19]. AIS has been used in other area of fields study, which are network engineering [19], data mining [20] and etc. Figure 3 show the basic clonal selection algorithm in AIS.

4. Web Mining Using Artificial Immune System (WMAIS)

In the Section 1, we have given a basic idea of WMAIS that help blog reader to find sources that relevant to them about political issues in Malaysia. In this section, we explain to you the basic idea of process flow in WMAIS based on the framework. As we have mentioned before, this WMAIS is inspired by AISIID but have been already renovated to be suited to Malay language so it can find relevant information about political issues in Malaysia through political weblog. Figure 4 show the framework of WMAIS.

From Figure 4, the process in WMAIS starts when a blog reader gives a seed URL and a set of keywords. The seed URL and keywords must be relevant to the political issues in Malaysia because it will help WMAIS to discover the information that has related to user needs. After that, the WMAIS will use the seed URL and keywords to learn, summaries and searching for relevant information on the Internet. While searching other relevant URL, the cells in AIS will help to organize the system by doing their speculation to measure the relevancy content of current URL, classify the relevant URLs and save it into database. After meet the user criterias, the cells will help to rank the relevant URLs based on the relevancy to the seed URL and present it to the blog reader.
In this research, the result from WMAIS will be evaluated by the blog reader. This is to make sure that the relevant information that have been searched by the WMAIS is fulfills the blog reader’s requirements. The blog reader needs to surf each URL and evaluate the content of each URL. Then the blog reader will be asked to give a rating to each of URL’s content based on their requirements. The blog reader needs to rank each URL’s content from 1 to 5 where 1 is for not relevant content and 5 for a content that is very relevant. The result of this rating will help us to calculate the result using a Student t-test tool. Student t-test is a statistical test and commonly used in data mining and machine learning texts [24]. This statistical tool is very suitable for experiment that has a small sample size. The result of calculation using Student t-test tool will show us whether the result of WMAIS is valid or not based on the user’s requirements.

![Figure 5. Example of social network of political weblog.](image)

The result also can help the blog reader discover other relevant information from the political weblog based on other weblog and each of web page from the weblog itself and other weblogs. In term of social network, it will show blog reader the social network of political weblog and how it can be related to each other. Indirectly, it will help on expanding the blog reader’s readership to find more relevant information about political issues in Malaysia. Figure 5 shows the example of social network of political weblog based on our observation on blogger readership in Malaysia.

5. DISCUSSION & CONCLUSION

In Malaysia, political blog has become one of the most powerful weapons used by the political party leaders to reach the voters. The usage of political blog gives authorities to the blog’s owner to give their opinions, spread information and sharing knowledge about political issues in Malaysia with other bloggers or readers. The existing of many political blogs on the Internet gives a wide readership and knowledge about political issues in Malaysia to blog readers. Sometimes, it will influence the blog readers who are turned out to be voters and these voters can make an impact to political situation in Malaysia. We build a web mining using Artificial Immune System (WMAIS) technique as search tool and to examine the relevant web pages and weblogs about political issues in Malaysia. The main thing in this research, we will stress on the development of WMAIS in order to find relevant information on political issues in Malaysia on the Internet.

6. ACKNOWLEDGMENTS

The authors wish to thank the reviewers for their helpful suggestions. Also thanks to all current and previous members of the Intelligence Software Engineering Lab (ISELab) at the Universiti Teknologi Malaysia (UTM). This work is supported by the Ministry of Science & Technology and Innovation (MOSTI), Malaysia and Research Management Centre, Universiti Teknologi Malaysia (UTM) under the Vot 79267.

7. REFERENCES


