

**FORMAL SPECIFICATION AND VALIDATION  
FOR PATTERN SCANNING**

**JULAILY AIDA BINTI JUSOH**

**MASTER OF SCIENCE  
UNIVERSITI MALAYSIA TERENGGANU  
MALAYSIA**

**2009**



**FORMAL SPECIFICATION AND VALIDATION  
FOR PATTERN SCANNING**

**JULAILY AIDA BINTI JUSOH**

**Thesis Submitted in Fulfilment of the Requirement for the  
Degree of Master of Science in the Faculty of Science and  
Technology  
Universiti Malaysia Terengganu  
November 2009**

Abstract of thesis presented to the Senate of Universiti Malaysia Terengganu  
in fulfilment of the requirement for the degree of Master of Science

**FORMAL SPECIFICATION AND VALIDATION FOR PATTERN SCANNING**

**JULAILY AIDA JUSOH**

**NOVEMBER 2009**

**Chairperson : Prof. Dr. Md. Yazid bin Mohd Saman, Ph.D**  
**Member : Mr. Mustafa Man**  
**Faculty : Science and Technology**

The requirements for a software system set out what the system should do and define constraints on its operation and implementation. Traditionally, users express the requirement specification of system development by natural language. Natural language is ambiguous, thus the requirement specification statement may result in different apprehension among users, analysts and programmers. Anxieties have been raised by several relevant software developers about the weakness use of natural languages in documenting system requirements. Hence, one of the solutions to solve the problem in capturing user requirements is by using the formal specification approach. Formal specifications can be utilized to provide an unambiguous and precise supplement to natural language descriptions. It can be rigorously validated, and verified leading to the early detection of specification errors.



In recent years, the number of DNA sequences increasing rapidly. Pattern scanning on large sequence database is one of the important functions in genomic research among biologist. It is useful for making new discoveries about a DNA sequence, including the location of functional sites and novel repetitive structures. Besides, it is also useful for the comparative analysis of different DNA sequences. Nevertheless, biologists face a problem in finding an accurate in pattern scanning in large DNA sequence database. The biologists frequently demand for optimal comparison result. So, the perfect structure of living beings evolution can be constructed. This task becomes more complex and challenging because the size of public sequence database is large and increasing exponentially every year.

Most studies in formal specifications have been carried out in critical and security areas. Nowadays, tools supporting formal methods are using *interactive theorem provers* and *model checkers*. Theorem proving is a part of validation techniques. Most software developers argue that the use of theorem proving technique in formal validation is not an essential part. This is because theorem proving technique for formal validation is only appropriate for complex applications. Nevertheless, the lack of formal support tools is a reason why it is not established among software developers.

This research will transform an informal specification for DNA sequences database system and pattern scanning into formal specification. Then both formal specification will be verify and validate using available support tools. In

this research, validation process on initial state theorem and pre-condition theorem done using theorem proving technique.

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

**SPESIFIKASI FORMAL DAN PENGESAHAN UNTUK  
PENGIMBASAN KATA KUNCI**

**JULAILY AIDA JUSOH**

**NOVEMBER 2009**

**Pengerusi : Prof. Dr. Md. Yazid bin Mohd Saman, Ph.D**  
**Ahli : Encik Mustafa Man**  
**Fakulti : Sains dan Teknologi**

Keperluan sistem perisian terdiri daripada apa yang perlu dilakukan oleh sistem dan mengenalpasti kekangan-kekangan yang terdapat pada operasi dan pelaksanaan system tersebut. Secara tradisi, para pengguna menyatakan keperluan spesifikasi pembangunan sistem dengan menggunakan bahasa tabii. Bahasa tabii bersifat kabur, oleh itu pernyataan spesifikasi keperluan boleh menimbulkan pelbagai tanggapan di kalangan pengguna, para penganalisis dan para pengaturcara. Kebimbangan telah ditimbulkan oleh beberapa pembangun perisian yang berkaitan mengenai kelemahan bahasa tabii yang sering digunakan dalam pendokumentasian keperluan sistem. Salah satu penyelesaian bagi menangani masalah dalam memahami keperluan pengguna adalah dengan menggunakan pendekatan spesifikasi formal. Spesifikasi formal adalah tidak kabur dan boleh digunakan sebagai kaedah yang jelas untuk huraian-huraian bahasa tabii.

Spesifikasi formal boleh disahsahkan bagi mengesan kesilapan yang ada pada peringkat awal.

Sejak kebelakangan ini, bilangan DNA bertambah dengan pesat. Pencarian persamaan DNA dalam suatu pangkalan data yang besar merupakan salah satu daripada fungsi penting dalam penyelidikan genomik di kalangan ahli biologi. Ia sangat berguna dalam penemuan DNA baru, termasuk lokasi tempat-tempat fungsian dan struktur yang berulang-ulang. Di samping itu, ia juga penting dalam membuat analisis bagi perbezaan yang terdapat pada DNA. Namun begitu, ahli biologi menghadapi masalah untuk mendapatkan persamaan yang tepat dalam pengimbasan kata kunci di dalam sesebuah pangkalan data DNA yang besar. Ahli biologi sering mengharapkan hasil perbandingan yang optimum. Jadi, kesempurnaan struktur kehidupan yang mengalami evolusi boleh didirikan. Tugas ini menjadi semakin kompleks dan mencabar apabila saiz pangkalan data yang sedia ada menjadi bertambah kepesatan pada setiap tahun.

Kebanyakan penyelidikan spesifikasi formal telah dijalankan bidang kritikal dan keselamatan. Kebelakangan ini, alat-alat yang menyokong kaedah-kaedah formal menggunakan *interactive theorem provers* dan *model checkers*. Pembuktian teorem merupakan sebahagian daripada teknik pengesahsahan. Kebanyakan pembangun perisian membantah bahawa penggunaan teknik pembuktian teorem yang terdapat dalam pengesahsahan secara formal adalah bukan bahagian yang penting. Ini kerana teknik pembuktian theorem dalam pengesahsahan secara formal lebih sesuai kepada aplikasi-aplikasi yang kompleks. Namun begitu,



kekurangan alatan sokongan bagi kaedah formal menyebabkan ia kurang diketahui dikalangan pembangun perisian.

Penyelidikan ini akan membangunkan spesifikasi formal dan pengesahsahihan bagi pangkalan data DNA dan pengimbasan kata kunci. Kemudian kedua-dua spesifikasi formal ini akan disahbetulkan dan disahsahihkan dengan menggunakan alat sokongan yang sedia ada. Dalam penyelidikan ini, proses pengesahsahihan ke atas teorem keadaan awal dan teorem pra-keadaan dijalankan dengan menggunakan teknik pembuktian teorem.