

แล้ว

การตรวจรู้ลายพิมพ์สีอื่นได้ด้วยคอมพิวเตอร์



นายกิตติศักดิ์ ตวีรัตน์

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
ภาควิชาวิศวกรรมคอมพิวเตอร์
บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย
พ.ศ. 2534
ISBN 974-578-668-3
ลิขสิทธิ์ของบัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย

017290 117225243

DNA FINGERPRINTS RECOGNITION BY COMPUTER

Mr. Kittisak Threerattananon

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science
Department of Computer Engineering
Graduate School
Chulalongkorn University

1991

ISBN 974-578-668-3



หัวข้อวิทยานิพนธ์ การตรวจรู้ลายพิมพ์ดีเอ็นเอด้วยคอมพิวเตอร์
โดย นายกิตติศักดิ์ ตรีรัตนนท์
ภาควิชา วิศวกรรมคอมพิวเตอร์
อาจารย์ที่ปรึกษา รองศาสตราจารย์ ดร. ศุภชัย ตั้งวงศ์ศานต์
 ผู้ช่วยศาสตราจารย์ สุยุษณ์ สัตยประกอบ

บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้หัวข้อวิทยานิพนธ์ฉบับนี้ เป็นส่วนหนึ่ง
ของการศึกษาตามหลักสูตรปริญญาโทบัณฑิต

.....
(ศาสตราจารย์ ดร. ถาวร วิชัยกร) คณบดีบัณฑิตวิทยาลัย

คณะกรรมการสอบวิทยานิพนธ์

.....
(ผู้ช่วยศาสตราจารย์ กอบกุล เตชะวิเศษ) ประธานกรรมการ

.....
(รองศาสตราจารย์ ดร. ศุภชัย ตั้งวงศ์ศานต์) อาจารย์ที่ปรึกษา

.....
(ผู้ช่วยศาสตราจารย์ สุยุษณ์ สัตยประกอบ) อาจารย์ที่ปรึกษาร่วม

.....
(อาจารย์ ดร. จิต ศิริบูรณ์) กรรมการ

พิมพ์ต้นฉบับบทคัดย่อวิทยานิพนธ์ภายในกรอบสี่เหลี่ยมนี้เพียงแผ่นเดียว

กิตติศักดิ์ ตรีรัตนนทร์ : การตรวจรู้ลายพิมพ์ดีเอ็นเอด้วยคอมพิวเตอร์ (DNA FINGERPRINTS RECOGNITION BY COMPUTER) อ.ที่ปรึกษา : รศ.ดร.ศุภชัย ตั้งวงศ์ศานต์, ผศ.สุยุชน์ สัตยประกอบ, 87 หน้า. ISBN 974-578-668-3

ดีเอ็นเอเป็นสิ่งที่ควบคุมการถ่ายทอดลักษณะทางกรรมพันธุ์ ในปัจจุบันได้มีการนำดีเอ็นเอมาผ่านขั้นตอนทางชีวเคมีทำให้เกิดเป็นแถบมืดสว่างบนแผ่นฟิล์ม เรียกว่าลายพิมพ์ดีเอ็นเอ (DNA fingerprints) จากการศึกษาพบว่าลายพิมพ์ดีเอ็นเอมีความเป็นเอกลักษณ์ และแตกต่างกันมากในแต่ละบุคคล สามารถนำมาใช้ชี้เฉพาะบุคคลหรือสืบหาความเกี่ยวข้องทางสายเลือด

ในการวิจัยนี้ ได้ทำการพัฒนาต้นแบบของโปรแกรมเพื่อการตรวจรู้ลายพิมพ์ดีเอ็นเอบนเครื่องไมโครคอมพิวเตอร์ประเภทไอพีเอ็มพีซี โดยมีเป้าหมายให้โปรแกรมตรวจรู้สามารถตรวจรู้รูปแบบของลายพิมพ์ดีเอ็นเอได้อย่างรวดเร็ว ถูกต้อง และสามารถตรวจหารูปแบบลายพิมพ์ที่ใกล้เคียงได้ เพื่อประโยชน์ในการตรวจสอบความเกี่ยวข้องทางสายเลือด โดยใช้โครงสร้างต้นไม้ค้นหาเชิงเลขฐานสอง (Binary digital search tree หรือ BDS-tree) อย่างกะชับ (compact) ในการแทนลายพิมพ์ดีเอ็นเอเพื่อการเก็บในคอมพิวเตอร์ รูปแบบของลายพิมพ์ดีเอ็นเอที่ต้องการตรวจรู้จะถูกป้อนเข้าสู่โปรแกรมผ่านทางแบ่นพิมพ์ ซึ่งมีลักษณะที่สำคัญคือ ตำแหน่งของลายพิมพ์, ความกว้างของลายพิมพ์ และความเข้มของลายพิมพ์ โดยจะใช้ตำแหน่งของลายพิมพ์ดีเอ็นเอเป็นคีย์ในการตรวจรู้ และเพื่อให้ตำแหน่งของลายพิมพ์เป็นมาตรฐานเดียวกัน ได้ใช้เทคนิคของการประมาณค่าด้วยโพลีโนเมียลตามวิธีลากรานจ์ (Lagrange polynomial interpolation) นอกจากนี้ได้จัดแบ่งกลุ่มของลายพิมพ์ดีเอ็นเอโดยใช้วิธีหาระยะเลเวนชไตน์ (Levenshtein distance)

ในการทดสอบต้นแบบของโปรแกรมตรวจรู้ลายพิมพ์ดีเอ็นเอ ได้ทดสอบจากคลังข้อมูลรูปแบบลายพิมพ์ดีเอ็นเอจำนวนประมาณ 50,000 รูปแบบ บนเครื่องไมโครคอมพิวเตอร์ประเภทไอพีเอ็มพีซี 386SX ผลปรากฏว่าสามารถตรวจรู้ได้ถูกต้องครบถ้วน ด้วยเวลาเฉลี่ย 1.03 วินาทีต่อหนึ่งรูปแบบลายพิมพ์ นอกจากนี้ยังสามารถค้นหารูปแบบลายพิมพ์ที่ใกล้เคียงที่มีความใกล้เคียงในระดับที่กำหนด ด้วยเวลาเฉลี่ย 4.47 วินาทีต่อหนึ่งรูปแบบลายพิมพ์



ภาควิชา วิศวกรรมคอมพิวเตอร์
สาขาวิชา วิทยาศาสตร์คอมพิวเตอร์
ปีการศึกษา 2533

ลายมือชื่อนิสิต
ลายมือชื่ออาจารย์ที่ปรึกษา
.....

พิมพ์ต้นฉบับภาคย่อยวิทยานิพนธ์ภายในกรอบสี่เหลี่ยมนี้เพียงแผ่นเดียว

KITTISAK THREEERATTANANON : DNA FINGERPRINTS RECOGNITION BY COMPUTER.
THESIS ADVISOR : ASSO. PROF.SUPACHAI TANGWONGSAN, Ph.D., ASST. PROF.
SUYUT SATAYAPRAKORB, 87 pp. ISBN 974-578-668-3

As it was discovered that hereditary transmission from one generation to another generation is controlled by DNA. By chemical process, DNA can be provided a set of genetic markers called DNA fingerprints which are completely specific to an individual and can be applied directly to problems of human identification or parenthood testing.

In this research, DNA fingerprints recognizer prototype is developed on an IBM PC microcomputer. As each individual DNA fingerprints pattern is uniquely characterized by the band position, width and intensity, the key approach here is to employ the compact binary digital search tree (BDS-tree) structure for pattern representation. For standard referencing of all DNA fingerprint patterns, the Lagrange polynomial interpolation is used to normalize the scale of pattern length. Furthermore, the prototype applies the technique of Levenshtein distance for pattern similarity measures in order to compare among DNA fingerprints. The recognizer after implementation is able to search not only fast and accurately, but also exhaustively search any similar patterns which can be used for parenthood testing.

The prototype is then tested under the database of approximate 50,000 patterns on a 386SX microcomputer. The result is that the recognizer performs the search for an existing pattern by an average time of 1.03 seconds. Not only the existing one, the recognizer can also search for 'close' patterns with a pre-defined level of similarity, the average time is 4.47 seconds per pattern.

ภาควิชา วิศวกรรมคอมพิวเตอร์
สาขาวิชา วิทยาศาสตร์คอมพิวเตอร์
ปีการศึกษา 2533

ลายมือชื่อนิสิต
ลายมือชื่ออาจารย์ที่ปรึกษา *Kit*



กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้ สำเร็จลุล่วงไปได้ด้วยความช่วยเหลือ ความร่วมมือ และคำแนะนำ อย่างดียิ่งจากท่านอาจารย์หลาย ๆ ท่าน ซึ่งผู้เขียนรู้สึกซาบซึ้งเป็นอย่างมาก จึงขอถือโอกาสนี้ กราบขอบพระคุณ

รองศาสตราจารย์ ดร. ศุภชัย ตั้งวงศ์ศานต์ และ ผู้ช่วยศาสตราจารย์ สุยุชน์ สัตยประกอบ อาจารย์ที่ปรึกษาวิทยานิพนธ์ ที่ได้มีเมตตาให้ความช่วยเหลือ คำแนะนำ และคำปรึกษา ที่เป็นประโยชน์อย่างยิ่งต่อการทำวิทยานิพนธ์

รองศาสตราจารย์ ดร. สกล พันธุ์ยิ้ม, รองศาสตราจารย์ ดร. วิชัย บุญแสง และอาจารย์ ชัยณรงค์ วงศ์ธีรทรัพย์ แห่งภาควิชาชีวเคมี คณะวิทยาศาสตร์ มหาวิทยาลัยมหิดล ที่ได้กรุณาให้ความรู้ ความเข้าใจ แก่ผู้เขียนในเรื่องของลายพิมพ์ดีเอ็นเอ

นอกจากนี้ ผู้เขียนขอขอบพระคุณ เจ้าหน้าที่ฝ่ายต่าง ๆ ของสำนักคอมพิวเตอร์ มหาวิทยาลัยมหิดล ที่ได้ให้ความร่วมมือ และอำนวยความสะดวกในการจัดเตรียมอุปกรณ์ บริภัณฑ์ต่าง ๆ เพื่อการพัฒนาและทดสอบระบบ รวมทั้งการจัดพิมพ์วิทยานิพนธ์ฉบับนี้

กิตติศักดิ์ ตรีรัตนหงษ์



สารบัญ

	หน้า
บทคัดย่อภาษาไทย	ง
บทคัดย่อภาษาอังกฤษ	จ
กิตติกรรมประกาศ	ฉ
สารบัญตาราง	ญ
สารบัญภาพและผังงาน	ฎ
บทที่	
1. บทนำ	1
1.1 ความเป็นมาของปัญหา	1
1.2 วัตถุประสงค์ของการวิจัย	1
1.3 ขอบเขตของการวิจัย	2
1.4 ขั้นตอนและวิธีดำเนินการวิจัย	2
2. การศึกษาค้นคว้าดีเอ็นเอและลายพิมพ์ดีเอ็นเอ	3
2.1 ประวัติการศึกษาดีเอ็นเอ	3
2.2 โครงสร้างของดีเอ็นเอ	4
2.3 ความหมายของลายพิมพ์ดีเอ็นเอ	6
2.4 การให้ได้มาซึ่งลายพิมพ์ดีเอ็นเอ	8
2.5 ประโยชน์ของลายพิมพ์ดีเอ็นเอ	11
2.5.1 การตรวจสอบความเกี่ยวข้องทางสายเลือด	11
2.5.2 การประยุกต์ใช้ในงานนิติเวชวิทยา	12
2.5.3 การตรวจสอบโรคมะเร็ง	12
2.5.4 การประยุกต์ใช้ในสัตว์	13
3. การออกแบบระบบการตรวจรู้ลายพิมพ์ดีเอ็นเอ	14
3.1 ลักษณะที่สำคัญของลายพิมพ์ดีเอ็นเอ	14
3.2 ขั้นตอนของการตรวจรู้ลายพิมพ์ดีเอ็นเอ	14
3.2.1 การแปลงมาตราส่วนตำแหน่งของลายพิมพ์ ให้เป็นมาตรฐานเดียวกัน	15

3.2.2	การจัดแบ่งกลุ่มของลายพิมพ์ดีเอ็นเอ	18
3.2.3	การตรวจรู้ลายพิมพ์ดีเอ็นเอจากตำแหน่ง ของลายพิมพ์ที่ใช้เป็นศิษย์	21
3.2.3.1	โครงสร้างต้นไม้มันหาเชิงเลขฐานสอง	21
3.2.3.2	การแทนต้นไม้มันบีตีสอย่างกระชับ	27
3.2.3.3	ตัวอย่างการแปลงตำแหน่ง ของลายพิมพ์เป็นศิษย์	31
3.2.4	การเปรียบเทียบหารูปแบบลายพิมพ์ที่ใกล้เคียง ...	31
4.	การพัฒนาระบบการตรวจรู้ลายพิมพ์ดีเอ็นเอ	34
4.1	เพิ่มข้อมูลที่ใช้ในการตรวจรู้ลายพิมพ์ดีเอ็นเอ	34
4.1.1	เพิ่มข้อมูลต้นไม้มันบีตีส	34
4.1.2	เพิ่มข้อมูลสมันยบิต	35
4.1.3	เพิ่มข้อมูลเลขที่อยู่ของถึง	35
4.1.4	เพิ่มข้อมูลรูปแบบลายพิมพ์ดีเอ็นเอ	36
4.2	ผังงานของการตรวจรู้ลายพิมพ์ดีเอ็นเอ	37
4.2.1	ผังงานแสดงขั้นตอนและวิธีดำเนินการ	37
4.2.2	ผังงานแสดงการทำงาน ของโปรแกรมตรวจรู้ลายพิมพ์ดีเอ็นเอ	37
4.3	การสร้างรูปแบบลายพิมพ์ดีเอ็นเอเพื่อการทดสอบ	38
5.	ผลการพัฒนาระบบการตรวจรู้ลายพิมพ์ดีเอ็นเอ	52
5.1	ข้อมูลเบื้องต้นเพื่อการทดสอบ โปรแกรมตรวจรู้ลายพิมพ์ดีเอ็นเอ	52
5.1.1	ระบบที่ใช้พัฒนารูปต้นแบบ	52
5.1.1.1	ระบบฮาร์ดแวร์	52
5.1.1.2	ระบบซอฟต์แวร์	52
5.1.2	โปรแกรมหลักและโปรแกรมยูนิต	52
5.1.3	ข้อมูลรูปแบบลายพิมพ์ดีเอ็นเอที่จะใช้ทำการทดสอบ	53
5.2	การทดสอบโปรแกรมตรวจรู้ลายพิมพ์ดีเอ็นเอ	53

6. สรุปผลการวิจัยและข้อเสนอแนะ	56
6.1 สรุปผลการวิจัย	56
6.1.1 ความถูกต้องของการตรวจรู้	56
6.1.2 ความเร็วของการตรวจรู้	56
6.1.3 การค้นหารูปแบบที่ใกล้เคียง	57
6.2 ข้อเสนอแนะ	59
6.2.1 การรับข้อมูลรูปแบบลายพิมพ์ดีเอ็นเอโดยตรง จากเครื่องกวาดตรวจ	59
6.2.2 การพัฒนาเป็นคลังข้อมูลสำหรับงานด้านอาชญากรรม	59
6.2.3 การใช้เทคนิคของการอัดข้อมูล	59
6.2.4 ไวยากรณ์ของการเรียงตัวของลายพิมพ์ดีเอ็นเอ ..	60
รายการอ้างอิง	61
ภาคผนวก	63
ประวัติผู้เขียน	87

สารบัญตาราง

ตารางที่	หน้า
3.1 แสดงน้ำหนักของตัวอักษรตามวิธีของ Bickel	32
5.1 (ก) ผลการทดสอบโปรแกรมตรวจรู้ กรณีที่ 1	54
5.2 (ข) ผลการทดสอบโปรแกรมตรวจรู้ กรณีที่ 2	55
5.3 (ค) ผลการทดสอบโปรแกรมตรวจรู้ กรณีที่ 3	55

สารบัญภาพและผังงาน

รูปที่		หน้า
2.1	แสดงนิวเคลียสไอโทดในดีเอ็นเอทั้ง 4 ชนิด	5
2.2	แสดงเกลียวคู่ตามแบบจำลองของวัตสันและคริก	5
2.3	แสดงการยึดคู่ระหว่างไซโตซีน-กวานีน และอะดีนีน-ไทมีน ด้วยพันธะไฮโดรเจน	7
2.4	แสดงลายพิมพ์ดีเอ็นเอของคน 9 คน โดยใช้เอนไซม์ตัดจำเพาะอินเอฟัน และดีเอ็นเอตรวจสอบของ Jeffreys และคณะ	9
2.5	แสดงลายพิมพ์ดีเอ็นเอของคน 9 คน โดยใช้เอนไซม์ตัดจำเพาะเฮย์ทรี และดีเอ็นเอตรวจสอบของ Vassart และคณะ	9
2.6	แสดงลายพิมพ์ดีเอ็นเอของฝาแฝดแท้ (7,8) เปรียบเทียบ กับของแม่ (6) และพ่อ (9)	9
2.7	แสดงขั้นตอนการสร้างลายพิมพ์ดีเอ็นเอ	10
3.1	แสดงขนาดมาตราส่วนมาตรฐาน ที่กำหนดขึ้น	17
3.2	แสดงวิธีของต้นไม้มอบีดีเอส	22
3.3	แสดงต้นไม้มอบีดีเอส จากรูปที่ 3.2 เมื่อมีการแยก	24
3.4	แสดงต้นไม้มอบีดีเอส จากรูปที่ 3.3 เมื่อมีการแยก	25
3.5 (ก)	แสดงต้นไม้มอบีดีเอส ก่อนการลบ	26
3.5 (ข)	แสดงต้นไม้มอบีดีเอส จากรูปที่ 3.5 (ก) หลังจากการลบ	26
3.5 (ค)	แสดงต้นไม้มอบีดีเอส เมื่อรวมกับถึงกลุ่มหมายเลข 2 จากรูปที่ 3.5 (ข)	27
3.6 (ก)	แสดงต้นไม้มอบีดีเอส	28
3.6 (ข)	แสดงตัวแทนแบบเชิงเส้นของต้นไม้มอบีดีเอส รูปที่ 3.6 (ก)	29
3.6 (ค)	แสดงตารางเลขที่อยู่ และสมนัยบิตของต้นไม้มอบีดีเอส รูปที่ 3.6 (ก)	29
3.7	แสดงมุมระหว่างเวกเตอร์ข้อคู่ต่าง ๆ	33
4.1	ลักษณะระเบียนของแฟ้มข้อมูลต้นไม้มอบีดีเอส	34
4.2	ลักษณะระเบียนของแฟ้มข้อมูลสมนัยบิต	35
4.3	ลักษณะระเบียนของแฟ้มข้อมูลเลขที่อยู่ของถึง	35
4.4	ลักษณะระเบียนของแฟ้มข้อมูลรูปแบบลายพิมพ์ดีเอ็นเอ	36
4.5	ขั้นตอนในการตรวจรู้ลายพิมพ์ดีเอ็นเอ	39

4.6	วิธีดำเนินการหาหมายเลขของถัง	40
4.7	วิธีดำเนินการค้นหารูปแบบลายพิมพ์ดีเอ็นเอ	41
4.8	วิธีดำเนินการเพิ่มข้อมูลรูปแบบลายพิมพ์ดีเอ็นเอ	42
4.9	วิธีดำเนินการค้นหารูปแบบลายพิมพ์ดีเอ็นเอที่ใกล้เคียง	43
4.10	การทำงานของโปรแกรมหลัก	44
4.11	การทำงานของโปรแกรมย่อย Input_Menu	45
4.12	การทำงานของโปรแกรมย่อย Process_Menu	46
4.13	การทำงานของโปรแกรมย่อย Options_Menu	47
4.14	การทำงานของโปรแกรมย่อย Go_Menu	48
4.15	การทำงานของโปรแกรมย่อย SearchKey	49
4.16	การทำงานของโปรแกรมย่อย DisplayInformation	50
4.17	วิธีดำเนินการสร้างรูปแบบลายพิมพ์ดีเอ็นเอเพื่อการทดสอบโดยการสุ่ม ..	51
6.1	แสดงกราฟของเวลาที่ใช้กับจำนวนรูปแบบลายพิมพ์ดีเอ็นเอ	58