

การใช้หลักคำพ้องเสียง เพื่อค้นหาชุดอักษรภาษาไทยที่ออกเสียงเหมือนกัน



นางสาวรวรณี อุกมพาศนิกย์

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

ภาควิชาวิศวกรรมคอมพิวเตอร์

บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย

พ.ศ. ๒๕๒๖

ISBN 974-562-163-3

011086

172119A3

STRING SEARCHING FOR THAI ALPHABET USING SOUNDEX COMPRESSION
TECHNIQUE

Miss Wannee Udompanich

A Thesis Submitted in Partial Fullfillment of the Requirements
for the Degree of Master of Science
Department of Computer Engineering
Graduate School
Chulalongkorn University
1983

หัวข้อวิทยานิพนธ์

โดย

ภาควิชา

อาจารย์ที่ปรึกษา

การใช้หลักคำพ้องเสียงเพื่อค้นหาชุดอักขระภาษาไทยที่ออกเสียงเหมือนกัน

นางสาววรรณิ อุกมพาดิษฐ์

วิศวกรรมคอมพิวเตอร์

นายสหัส ทวีทิพยบุตร



บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย ออนุมัติให้บัณฑิตวิทยาลัยนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรบัณฑิต

..... *[Signature]* คณะที่บัณฑิตวิทยาลัย
(รองศาสตราจารย์ ทวี สุประสิทธิ์ บูณาค)

คณะกรรมการสอบวิทยานิพนธ์

..... *[Signature]* ประธานกรรมการ
(รองศาสตราจารย์ เกื้อน สินธุพันธ์ประทุม)

..... *[Signature]* กรรมการ
(นายสหัส ทวีทิพยบุตร)

..... *[Signature]* กรรมการ
(รองศาสตราจารย์ ไกรวิชิต คันทิเมฆ)

..... *[Signature]* กรรมการ
(ผู้ช่วยศาสตราจารย์ ชูชีพ จิมวงษ์)

ลิขสิทธิ์ของบัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย

การวิจัยนี้ได้ใช้ข้อมูลซึ่งเป็นรายชื่อนิสิตทั้งสิ้น 2,925 ชื่อเป็นตัวอย่างทดสอบในการทดสอบได้สร้างแฟ้มข้อมูลจากตัวอย่างข้างต้นโดยมีรหัสค้นหาที่สร้างขึ้นด้วยกรรมวิธีดังกล่าวเป็น 2 ระบบเพื่อการเปรียบเทียบ และจากการทดสอบโดยใช้ชื่อเข้าค้นหาข้อมูลในแฟ้มข้อมูลทั้ง 2 ระบบ 220 ครั้ง ด้วยขนาดความยาวของรหัสจาก 3 ถึง 10 ปรากฏว่า ระบบที่ 2 ที่ขนาดความยาวของรหัสเป็น 7 ให้ประสิทธิภาพและประสิทธิผลที่สุด

Thesis Title String Searching for Thai Alphabet
 Using SOUNDEX Compression Technique
Name Miss Wannee Udompanich
Thesis Advisor Mr. Sahas Treetipbut
Department Computer Engineering
Academic Year 1982



ABSTRACT

In the computer personnel information retrieval system, the identification number is normally used as a primary key for searching data. Actually this way is not convenient for users to use because users have to remember or know the identification number. However if personnel name is used as a primary key instead, several problems will consequently occur. Users will face these problems when they are provided with wrong spelling records or incorrect-sequential-ordered records.

Many techniques have been developed and researched in order to facilitate users to retrieve correct data with most efficiency and productivity by the way of using name as a key. One of the most famous algorithms is to transform the name wanted to search to a special code by applying English SOUNDEX Methodology. The principle of identical sound of the first Thai alphabet and the English SOUNDEX

Methodology are applied together in order to search the name written in Thai more easily and efficiently because Thai and English are very much different in grammatical rules and written-styles.

The treatment employs 2,925 student names for sample testing. The sample file is created by applying the two searching code systems for comparison. In order to compare the performance of both systems, 220 searchings are trialed each system. The 220 experiments employ the key length of characters varying from 3 to 10. The result shows that the second system with a key length of seven characters gives the best performance.



กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้ สำเร็จได้ก็ด้วยความช่วยเหลือจาก อาจารย์ สหีส ตรีทิพย์บุตร อาจารย์ผู้ควบคุมการทำวิทยานิพนธ์ รองศาสตราจารย์ ไกรวิชิต กันติเมธ อาจารย์ผู้ควบคุมการทำวิทยานิพนธ์ร่วม ที่ได้กรุณาแนะนำหัวข้อวิทยานิพนธ์ และแนวทางการทำวิจัย การเขียน และตรวจแก้ไขตั้งแต่ต้นจนสำเร็จเป็นเล่ม ผู้เขียนขอกราบขอบพระคุณท่านอาจารย์ทั้งสองเป็นอย่างสูง

นอกจากนี้ก็ขอขอบพระคุณคณะกรรมการสอบวิทยานิพนธ์ และเจ้าหน้าที่ทะเบียนกลาง จุฬาลงกรณ์มหาวิทยาลัย ที่ได้อำนวยความสะดวกแก่ผู้เขียนในการคัดลอกรายชื่อนิสิต เพื่อนำมาใช้เป็นข้อมูลในการทำวิจัย ไว้ ณ ที่นี้ด้วย

วรรณิ อุกมพาดิษฐ์



สารบัญ

	หน้า
บทคัดย่อภาษาไทย	ง
บทคัดย่อภาษาอังกฤษ	ฉ
กิตติกรรมประกาศ	ช
รายการตารางประกอบ	ฉ
รายการรูปประกอบ	ฉ
บทที่	
1 บทนำ	1
1.1 ความเป็นมาของปัญหา	1
1.2 วัตถุประสงค์ของการวิจัย	2
1.3 ขอบเขตของการวิจัย	3
1.4 ขั้นตอนในการทำวิจัย	3
1.5 ประโยชน์ที่ได้รับจากการวิจัย	4
2 ชาวคึกคัก	5
2.1 ความหมายของชาวคึกคัก	5
2.2 การกักแปลงไปชาวคึกคักกับภาษาไทยและปัญหา	8
2.3 การจกกลุ่มพยัญชนะไทย	13
2.4 หลักเกณฑ์และข้อกำหนดต่าง ๆ ในการให้รหัส	18
3 การพิจารณาการจกเพิ่มข้อมูลหรือวิธีปฏิบัติข้อมูล	21
3.1 การจกเพิ่มข้อมูลแบบค่อเป็อง	22
3.2 การจกเพิ่มข้อมูลแบบสุ่ม	26

สารบัญ (ต่อ)

บทที่	หน้า
3.3 การจักเพิ่มข้อมูลแบบต่อเนื่องมีคชันี	31
3.4 พิจารณาเปรียบเทียบการจักเพิ่มข้อมูล	35
4 ใช้โปรแกรมทดสอบเพื่อหาขนาดของรหัสที่เหมาะสม	38
4.1 การทดสอบและผลจากการทดสอบ	39
4.2 พิจารณาขนาดของรหัสที่เหมาะสม	47
4.3 เปรียบเทียบการจักกลุ่มที่ 1 และการจักกลุ่มที่ 2	55
5 สรุปผลการวิจัยและขอเสนอแนะ	57
5.1 สรุปผลการวิจัย	57
5.2 ขอเสนอแนะ	58
บรรณานุกรม	63
ภาคผนวก	65
ประวัติ	85

รายการตารางประกอบ

ตารางที่		หน้า
4.1	โอกาสที่ค้นหาข้อมูลไม่พบ โอกาสที่ค้นหาข้อมูลได้ 1 ชื่อและถูกต้อง และโอกาสที่ชื่อต่างกันได้รหัสซ้ำกัน จากการค้นหาข้อมูลโดยอาศัยวิธีการ จัดกลุ่มที่ 1	47
4.2	โอกาสที่ค้นหาข้อมูลไม่พบ โอกาสที่ค้นหาข้อมูลได้ 1 ชื่อและถูกต้อง และโอกาสที่ชื่อต่างกันได้รหัสซ้ำกัน จากการค้นหาข้อมูลโดยอาศัยวิธีการ จัดกลุ่มที่ 2	51

รายการรูปประกอบ

รูปที่		หน้า
3.1	แสดงเพิ่มข้อมูลแบบต่อเนื่องและวิธีค้นหา	25
3.2	แสดงเพิ่มข้อมูลแบบสุ่มและวิธีค้นหา	29
3.3	แสดงเพิ่มข้อมูลแบบต่อเนื่องมีคีย์และวิธีค้นหา	34
4.1 ก	โอกาสที่ค้นหาข้อมูลไม่พบของการจัดกลุ่มที่ 1	48
4.1 ข	โอกาสที่ค้นหาข้อมูลได้ 1 ชื่อและถูกต้องของการจัดกลุ่มที่ 1 ..	49
4.2 ก	โอกาสที่ค้นหาข้อมูลไม่พบของการจัดกลุ่มที่ 2	52
4.2 ข	โอกาสที่ค้นหาข้อมูลได้ 1 ชื่อและถูกต้องของการจัดกลุ่มที่ 2 ..	53
4.3	แสดงการเปรียบเทียบการจัดกลุ่มที่ 1 และการจัดกลุ่มที่ 2 ...	55
5.2 ก	การสร้างเพิ่มข้อมูลแบบต่อเนื่องมีคีย์	60
5.2 ข	ตัวอย่างการค้นหาข้อมูลจากเพิ่มข้อมูลแบบต่อเนื่องมีคีย์	61