

แบบเสนอโครงการวิจัยพัฒนา  
“ระบบการค้นหาคำภาษาไทยแบบ Full-Text Search บน MySQL”  
สถาบันส่งเสริมการจัดการความรู้เพื่อสังคม

---

1. ชื่อโครงการ

Thai Full-Text Search System on MySQL  
ระบบการค้นหาคำภาษาไทยแบบ Full-Text Search บน MySQL

2. ชื่อนักพัฒนา หน่วยงานที่สังกัด ที่อยู่ หมายเลขโทรศัพท์ โทรสาร และ E-mail

ดร.รัชชัช ปิยะวัฒน์  
คณะวิทยาการจัดการ มหาวิทยาลัยสงขลานครินทร์  
ตู้ ป.ณ. 5 คอหงษ์ อ.หาดใหญ่ จ. สงขลา 90112  
โทรศัพท์: 0-7428-7934 โทรศัพท์มือถือ: 0-9737-9137  
โทรสาร: 0-7421-2819 E-mail: piyawat@usablelabs.com

ดร.จันทวรรณ น้อยวัน  
คณะวิทยาการจัดการ มหาวิทยาลัยสงขลานครินทร์  
ตู้ ป.ณ. 5 คอหงษ์ อ.หาดใหญ่ จ. สงขลา 90112  
โทรศัพท์: 0-7428-7931 โทรศัพท์มือถือ: 0-9649-6995  
โทรสาร: 0-7421-2819 E-mail: noiwan@usablelabs.com

3. ระยะเวลาการศึกษาวิจัย

ปี พศ. 2549 - 2550

4. หลักการและเหตุผล

ภาษาไทยนับเป็นหนึ่งในภาษาที่ซับซ้อนของโลกเพราะด้วยคำที่ประกอบด้วยทั้งตัวอักษรสระ และวรรณยุกต์ ที่เรียงติดต่อกันและลักษณะการเขียนที่ไม่มีการเว้นช่องว่างระหว่างคำของภาษาไทย ทำให้การตัดคำ (Word Boundary Segmentation) มีโอกาสเกิดความกำกวม ต้องอาศัยกระบวนการวิธี (Algorithm) ที่พัฒนาขึ้นมาเฉพาะ ไม่สามารถทำได้โดยง่ายดังภาษาที่ใช้ช่องว่างในการแบ่งคำเช่นภาษาอังกฤษ

ด้วยเหตุนี้ การพัฒนาระบบสารสนเทศหรือซอฟต์แวร์ที่ต้องมีการจัดการเกี่ยวกับภาษาไทย จึงมีความยากลำบากกว่าการพัฒนาที่ใช้ภาษาอังกฤษเพียงภาษาเดียว

นอกจากนี้ในการจะนำซอฟต์แวร์ที่พัฒนาขึ้นเพื่อภาษาอังกฤษมาใช้กับภาษาไทยได้นั้น ต้องอาศัยการดัดแปลงเพื่อให้สามารถทำงานได้กับลักษณะเฉพาะดังกล่าวของภาษาไทยเสียก่อน จึงสามารถใช้ซอฟต์แวร์นั้นได้อย่างเต็มประสิทธิภาพ

เนื่องจากลักษณะเฉพาะของภาษาไทยดังกล่าว ทำให้การพัฒนาระบบเพื่อกลุ่มผู้ใช้คนไทย โดยเฉพาะอย่างยิ่งระบบที่ต้องมีการจัดเก็บข้อมูลต่างๆ ลงในฐานข้อมูลมีความยากลำบากประการสำคัญ คือการเพิ่มขีดความสามารถของระบบในการค้นหาคำภาษาไทย

ในปัจจุบัน ระบบการจัดการฐานข้อมูลแบบ Open-Source ที่เป็นที่นิยมใช้กันอย่างแพร่หลายในประเทศไทยและในระดับโลก คือ MySQL ซึ่ง MySQL ตั้งแต่เวอร์ชัน 3.23.23 เป็นต้นมา มีความสามารถที่มีประโยชน์อย่างยิ่งอย่างหนึ่ง คือ การค้นหาแบบ Full-Text (Full-Text Search) การค้นหาแบบนี้จะแตกต่างจากการค้นหาแบบเปรียบเทียบคำ (String Comparison) โดยการค้นหาแบบ Full-Text สามารถค้นหาคำที่ใกล้เคียงกับคำที่ค้นหาได้ เรียกว่า Blind Query Expansion หรือ Automatic Relevance Feedback

ตัวอย่างเช่น หากผู้ใช้ค้นหาคำว่า “Database” ด้วย Full-Text Search ผลจากการค้นหาจะพบเอกสารที่ประกอบด้วยคำว่า “MySQL” “DBMS” หรือ “Oracle” เป็นต้นด้วย เพราะว่า คำเหล่านี้เป็นคำที่สอดคล้องเกี่ยวข้องกับคำว่า “Database” ทั้งสิ้น หรือในอีกกรณีหนึ่ง หากผู้ใช้ไม่มั่นใจในความถูกต้องของการสะกดคำที่ค้นหา เช่น ชื่อบุคคล ด้วยความสามารถของการค้นหาแบบ Full-Text Search การค้นหาชื่ออื่นๆ และคำประกอบที่เกี่ยวข้องจะทำให้การค้นหาที่น่าสนใจไปสู่ผลลัพธ์ที่ผู้ใช้ต้องการจริงๆ ได้

แต่อย่างไรก็ตาม ในปัจจุบันนี้ การค้นหาคำภาษาไทยด้วย Full-Text Search ยังไม่ได้รับการพัฒนา อันอาจเนื่องมาจากความยากในการประมวลผลของภาษาไทยและการขาดแคลนโปรแกรมเมอร์คนไทยที่มีความชำนาญในภาษา C (ซึ่งเป็นภาษาที่ใช้ในการพัฒนา MySQL) เพื่อทำการแก้ไข MySQL ให้สามารถทำงานแบบ Full-Text Search กับภาษาไทยได้

ในขณะเดียวกัน ความต้องการในการใช้งาน Full-Text Search ภาษาไทยนั้นก็กลับมีเพิ่มสูงขึ้นอย่างรวดเร็วอันเนื่องมาจากการเติบโตของระบบรูปแบบใหม่ๆ อาทิเช่น ระบบบล็อก (Blog), ระบบเขียนข้อความร่วมกัน (Wiki) และระบบการจัดการเนื้อหาข้อมูล (Content Management System)

## 5. วัตถุประสงค์

- 5.1. เพื่อเพิ่มศักยภาพในการค้นหาคำภาษาไทยแบบ Full-Text Search ใน MySQL
- 5.2. เพื่อเผยแพร่ MySQL ที่มีความสามารถในการประมวลผลภาษาไทยแบบ Full-Text Search ให้ใช้งานอย่างแพร่หลาย
- 5.3. เพื่อเป็นตัวอย่างต้นแบบและจุดเริ่มต้นเพื่อถ่ายทอดความรู้ฝังลึกในการพัฒนาซอฟต์แวร์แบบ Open-Source ให้รองรับภาษาไทยได้อย่างถูกต้องและเต็มประสิทธิภาพของซอฟต์แวร์

## 6. กลุ่มเป้าหมาย

- 6.1. กลุ่มผู้พัฒนาซอฟต์แวร์เพื่อให้ใช้งานได้กับภาษาไทย ทั้งซอฟต์แวร์บนเครื่องลูก (Desktop Applications) ซอฟต์แวร์บนเครื่องแม่ข่าย (Server-Based Applications) รวมถึงซอฟต์แวร์ที่ทำงานผ่านเว็บ (Web-Based Applications)
- 6.2. กลุ่มผู้พัฒนาระบบการจัดการเนื้อหาข้อมูล (Content Management System) หรือระบบบล็อก (Blog) หรือ ระบบเขียนข้อความร่วมกัน (Wiki) เพื่อคนไทย

## 7. ขอบเขตการพัฒนา

- 7.1. พัฒนาระบบการค้นหาคำภาษาไทยแบบ Full-Text (Full-Text Search) เพื่อให้ทำงานบนระบบการจัดการฐานข้อมูล MySQL
- 7.2. เปิดเผยแพร่ Source Code (Open-Source)
- 7.3. เผยแพร่ Binary Versions ของ “MyThaiSQL” (MySQL ที่ปรับปรุงกับภาษาไทยแล้ว) ให้นักพัฒนาสามารถนำไปใช้งานได้ ทั้งที่ทำงานได้บน Microsoft Windows, Linux, และ FreeBSD
- 7.4. นำเสนอ Source Code ส่วนที่เพิ่มเติมนั้นให้แก่ MySQL AB (บริษัทผู้ผลิต MySQL) เพื่อรวมเข้าเป็นส่วนหนึ่งของ MySQL

## 8. กำหนดการ

เป็นเวลา 12 เดือน นับตั้งแต่วันที่ลงนามในสัญญาจ้าง ประมาณการกำหนดการดังนี้คือ

ตุลาคม 2548	นำเสนอโครงการ และ ลงนามในสัญญาจ้าง
พฤศจิกายน 2548	ศึกษาข้อมูล
มกราคม 2549	พัฒนาระบบ
ตุลาคม 2549	เสร็จสิ้นการพัฒนาและพร้อมใช้งาน

**9. งบประมาณ**

งบประมาณรวมทั้งสิ้น 200,004 บาท แบ่งเป็น

9.1. ค่าจ้างทีมงานนักพัฒนา

9.2. ค่าอุปกรณ์เทคโนโลยี และ

9.3. ค่าแหล่งข้อมูลเพื่อการศึกษาพัฒนา เช่น หนังสือ เป็นต้น

โดยสัดส่วนค่าใช้จ่ายในแต่ละส่วนนั้นให้อยู่ในดุลพินิจของหัวหน้าโครงการ โดยมี  
กำหนดการจ่ายเงินเดือนละ 16,667 บาท ทุกเดือนเป็นเวลา 12 เดือน