



ThaiEngine Innovative



Full-text search

Sound Like/Synonym

Partial Search

Match Highlight

Suggest Search Keyword

SQL like Interface

Relevance Ranking

Statistic Report

ThaiEngine เป็นทั้ง Information Retrieval System และ Indexer โดยผู้ใช้จะค้นหาข้อมูลโดยการสื่อสารกับโปรแกรมทั่วไปที่ทำหน้าที่เป็น User Interface System (UI System) ซึ่งจะค้นหาข้อมูลผ่าน IR System และส่งผลลัพธ์ไปให้ผู้ใช้ในรูปแบบที่ผู้ใช้เข้าใจได้ง่าย ในขณะที่ตัว IR System จะทำงานกับฐานข้อมูลของตัวเอง ซึ่งเป็นผลผลิตจากการอ่านและวิเคราะห์ข้อมูลของ Indexer

การค้นหาข้อมูลใน ThaiEngine

- ทำงานได้ดีกับข้อมูลที่นำมาจากระบบฐานข้อมูลทั่วไป ซึ่งมักมีลักษณะสำคัญคือ มีลักษณะเป็น row, column มีปริมาณมาก และมักไม่มีลักษณะเป็นประโยคภาษาไทยที่สมบูรณ์
- ทำงานโดยใช้พยางค์เป็นหน่วยย่อยสุด ทำให้ผู้ใช้สามารถค้นหาคำในลักษณะพิเศษได้ เช่น ค้นหาด้วยข้อความ "ใบไม้เขียว" ก็จะพบข้อมูลที่มี "ใบ" "ไม้" และ "เขียว" บรรจุอยู่ในลักษณะที่สลับกันได้ เช่น พบ "ใบไม้สีเขียว" , "สีเขียวในใบไม้" และ "ต้นไม้ใบเขียว" ด้วย หรือหากค้นหาด้วยข้อความ "baimai" ก็จะพบข้อมูลที่มี "bai mai" และ "mai bai" ด้วย เป็นต้น
- สามารถบีบผลการค้นหาให้มีจำนวนน้อยลงโดยใช้คุณสมบัติ exact match คือ ต้องมีข้อความที่ระบุเท่านั้น ไม่เอากการสลับพยางค์ หรือเสียงพ้อง
- สามารถบีบผลการค้นหาให้มีจำนวนน้อยลง ในลักษณะพิเศษ (คล้ายกับ regular expression) ได้ เช่น ต้องการให้มีเฉพาะข้อมูลที่อยู่ในรูป "ใบไม้*เขียว*อาหาร" โดยที่ * หมายถึงพยางค์ใดๆ ตั้งแต่ 0 พยางค์ขึ้นไป
- ThaiEngine รู้จักข้อมูลในรูปแบบอื่นที่ไม่ใช่ text ด้วย เช่น numeric, date, time, date+time เป็นต้น ทำให้ผู้ใช้สามารถระบุลักษณะของข้อมูลที่ต้องการเพิ่มเติมบางอย่างได้ เช่น ต้องการเฉพาะข้อมูลที่มีราคามากกว่า 1,000 ขึ้นไป
- มีระบบการเก็บผลการค้นหาไว้ใน cache memory ทำให้สามารถดูผลการค้นหาหน้าต่อไป หรือหน้าก่อนได้อย่างรวดเร็ว ไม่ต้องทำการค้นหาใหม่ ทำให้ระบบสามารถรองรับปริมาณการค้นหาได้มากขึ้น
- มีความเร็วในการค้นหาสูง response time ของการค้นหาทั่วไปอยู่ที่ 0.01 ถึง 0.2 วินาที และโดยรวมรองรับได้ 1,000 ถึง 2,000 search ต่อวินาที บนเครื่อง Pentium 4, 1 processor, 2.0 GHz, 2GB RAM ทั่วไป

FEATURES

Full-text search:

ทำงานได้ดีกับข้อมูลที่ได้จากระบบฐานข้อมูลโดยเฉพาะ เป็นระบบ 2 ภาษาคือ ไทย-อังกฤษ ใช้พยางค์เป็นหน่วยย่อยที่สุดในการประมวลผล การออกแบบใช้สมมติฐานดังนี้

- ข้อมูลที่ใช้อักขระไทย อาจเป็นได้ทั้งภาษาไทย ภาษาอังกฤษ หรือภาษาอื่นก็ได้
- ข้อมูลที่ใช้อักขระลาติน อาจเป็นภาษาอังกฤษ ภาษาอเมริกัน ภาษาไทย หรือภาษาอื่นๆ ก็ได้
- คำ/พยางค์ไทย ที่เราพบในข้อมูล อาจพบอยู่ในพจนานุกรมไทย หรือไม่ก็ได้
- คำ/พยางค์ภาษาอังกฤษ อาจพบหรือไม่พบในพจนานุกรมอังกฤษ/อเมริกันก็ได้

มีระบบการตัดพยางค์ เช่น ตัดพยางค์ -> ตัด-พ-ยางค์ หรือ samudprakan -> sa-mud-pra-kan จะเห็นว่า ThaiEngine สามารถตัดพยางค์ภาษาอังกฤษได้ด้วย สามารถค้นหาข้อมูลต่างๆ ได้อย่างเป็นธรรมชาติ เช่น หา "ใบไม้" จะพบ "ใบไม้" , "ไม้ใบ" , "ไม้xxxใบ" แต่จะไม่พบ "ใบไม้" หรือ หา "samud" จะพบ "samudprakan" "samudsongkram" และสามารถบีบให้เอาผลลัพธ์เฉพาะที่เป็น exact match ได้ เช่น จากตัวอย่าง "ใบไม้" สามารถระบุว่ามี "ไม้ใบ" , "ไม้xxxใบ" ได้

จะเห็นว่าทำงานได้ดีกับ คำที่ไม่พบในพจนานุกรม(ไทย-อังกฤษ)ซึ่งมีส่วนสูงมากในข้อมูลภาษาไทยในปัจจุบัน



Nysiis Solutions Co., Ltd.

159/6 Serm-Mit Tower, 2nd Floor, Sukhumvit 21 Road (Asoke), North-Klongtoey Sub-District, Wattana District, Bangkok 10110
 T: +66 (2) 665 7157-8
 F: +66 (2) 665 7159
info@nysiissolutions.com
www.nysiissolutions.com



- Full-text search
- Sound Like/Synonym
- Partial Search
- Match Highlight
- Suggest Search Keyword
- SQL like Interface
- Relevance Ranking
- Statistic Report



Support Platform:

- Solaris 10
 - SPARC or X86
 - RAM 4GB,HDD 100GB
- Linux – Red hat, Ubuntu
 - X86
 - RAM 4GB,HDD 100GB

Sound Like/Synonym:

Sound Like มีระบบการแปลงพยางค์ทั้งภาษาไทยและภาษาอังกฤษ ไปเป็นรหัสเสียงแบบ phonetic (ซึ่งเป็นงานวิจัยเฉพาะของผู้ผลิต Thai Engine) ผู้ใช้จึงสามารถค้น หาแบบพ้องเสียงได้หลายรูปแบบ ที่ไม่เคยปรากฏมาก่อนในวงการ IT ไทย เช่น

- ค้น "สมศักดิ์" จะพบ "สมสัก" "สมศักดิ์" "สมศักดิ์" "somsak" "ซมซั๊ก"
- ค้น "ไบไม" จะพบ "ไบไม" "ไบไม" "บัยมัย" "bimi" "bimai" "baimi" "bi mai"

หมายเหตุ ThaiEngine ใช้ text-to-phonetic ไม่ใช่ soundex

Synonym หมายถึง กลุ่มของ "วลี" ที่มีความหมายเหมือนกัน เช่น สถานีตำรวจ = สน. , โรงเรียน = รร. , บ้าน = เรือน, house, home

- ผู้ใช้สามารถค้นหาโดยให้ระบบนำเอาข้อมูลที่เป็นคำพ้องออกมาได้ด้วยได้
- Thai Engine มีฐานข้อมูลคำเหมือนต่างๆ ไว้ให้ผู้ใช้เลือก เช่น คำพ้องจาก พจนานุกรม อังกฤษ-อังกฤษ ไทย-ไทย ไทย-อังกฤษ ผู้ใช้สามารถเพิ่มเติมข้อมูลคำพ้องภายในองค์กรเข้าไปได้เอง

หมายเหตุ ThaiEngine ไม่ได้ใช้ stem algorithm

Partial Search:

ระบบการค้นหาแบบส่วนหนึ่งของพยางค์/คำ ในข้อมูลบางอย่าง เช่น หมายเลขโทรศัพท์(0816996269, 029504381) หรือ รหัสสินค้า เช่น (ABC9383, 94883048) ซึ่งถือเป็น 1 พยางค์ ปกติแล้วผู้ใช้จะไม่สามารถกรอกรายละเอียดได้ทุกตัวอักษร

- Thai Engine มีระบบการทำ index แบบพิเศษซึ่งสามารถทำสร้าง index แบบส่วนหนึ่งของพยางค์ได้ เช่น นำเอาตัวอักษรใดๆ ที่อยู่ติดกัน ตั้งแต่ 4 ตัวอักษร ถึง 7 ตัวอักษรไปทำ index ด้วย ซึ่งในกรณี 0816996269 ผู้ใช้จะสามารถค้นพบข้อมูลด้วย keyword ดังนี้ 0816, 1699, 6269, 81699, ..., 6996269

Match Highlight:

Match Highlight ระบบสามารถ เน้นข้อความที่ตรงกับ search keyword ได้ การเน้นสามารถปรับเปลี่ยนได้ตามต้องการ เช่น

- เน้นด้วย คู่ "[" และ "]" , "
" และ "</br>"
- เน้นด้วย คู่ "" และ "" 2550

Suggest Search Keyword:

ระบบแนะนำ Keyword ในการค้นหาให้กับผู้ใช้ โดยแสดงคำแนะนำที่ใกล้เคียงกับข้อความค้นหามาแสดง เพื่อให้ผู้ใช้สามารถใช้งานหรือค้นหาข้อมูลได้สะดวกยิ่งขึ้น

SQL like Interface:

ThaiEngine ใช้คำสั่งที่คล้ายกับ SQL ในการติดต่อระหว่าง programmer กับ engine ทำให้สามารถปรับแต่ง application ได้ง่าย support search stored procedure

Relevance Ranking:

ThaiEngine สามารถการควบคุมลำดับของข้อมูลที่ส่งให้ผู้ใช้ได้ทั้งแบบที่เรียงตาม column ที่ต้องการ หรือเรียงตามคะแนนความสอดคล้อง (page rank) Relevance Ranking ของ ThaiEngine สามารถปรับเปลี่ยนได้หลายแบบ ขึ้นอยู่กับข้อกำหนดร่วมกันระหว่าง ThaiEngine และผู้พัฒนาระบบ application

Statistic Report:

ThaiEngine มีรายงานแสดงสถิติการใช้ระบบไว้จำนวนหนึ่งเพื่อให้เจ้าของระบบสามารถตรวจสอบพฤติกรรมการใช้งานของผู้ใช้ ดังนี้

- Daily System Statistic Report
- Search Report Classified by Result Rows Report
- Monthly System Statistic Report
- Search Log XML Report