**Overview**

Diagram

Description automatically generated

**หน้าที่ของแต่ละ Services**

**Web service back-end**

ทำหน้าที่เป็น web api สำหรับให้ User ส่งข้อมูลการค้นหาเข้ามาในระบบพร้อมกับทำหน้าทีจัด Format โดยจะทำการเรียกใช้ Service อีก 2 ตัวเพื่อใช้ในการทำ Format สำหรับส่งให้ Search Engine เพื่อค้นหาข้อมูลสถานที่และนำผลลัพธ์ที่ได้ส่งกลับไปยัง User

**Word break service**

Service สำหรับการตัดคำ โดยจะมีตัว Dictionary เพื่อใช้ในการตัดคำ เช่น สนามบินดอนเมือง เมื่อเข้าทำงานที่ Service ก็จะทำการ Word break ออกมาเป็น สนาม บิน ดอน เมือง เเล้วนำ Word Break ที่ได้ออกมา ไปเข้า Service Soundex ต่อไป

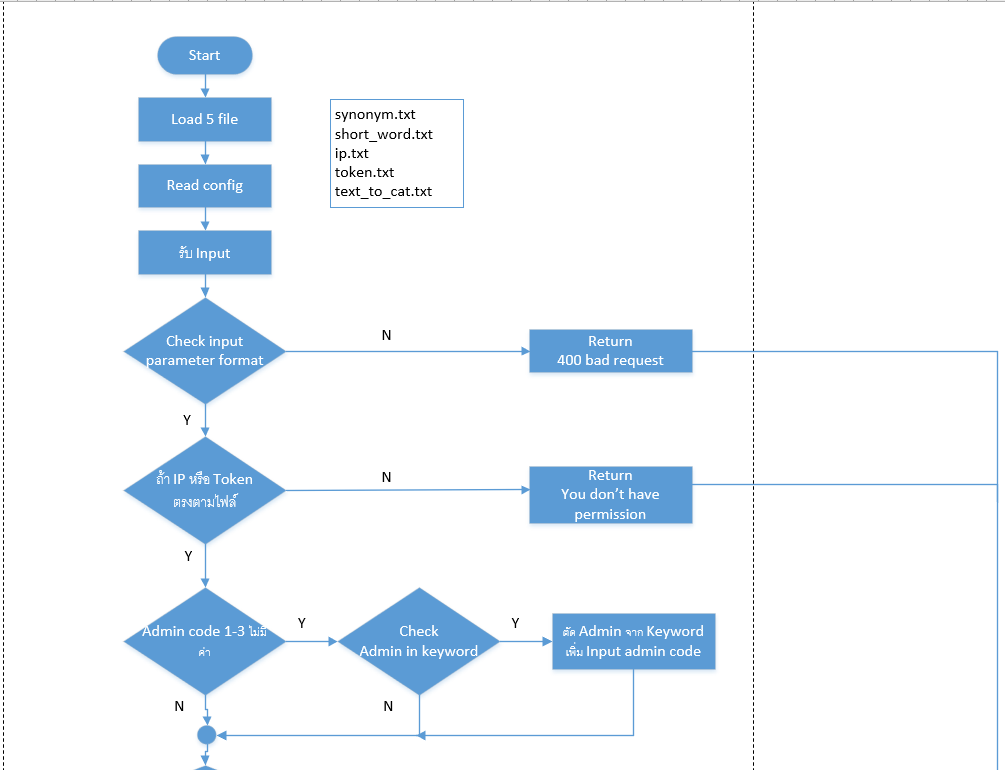
**Soundex service**

Service สำหรับแปลง Keyword เป็น soundex และจำเป็นต้องมีชุด Data เพื่อใช้ในการทำ soundex จึงจะได้เป็นชุดอักษรเสียงมา

**Search Engine**

ใช้สำหรับเป็น Search Enine ทำงานแบบ Websocket Engine Search จะคอยรับ message หรือ String format มาจาก Web Service Search backend จากนั้นนำไปค้นหาและส่งผลการคำนวณกลับไปยัง Web Service Search backendในการรัน Engine Search จำเป็นต้องมีชุดข้อมูล Index ด้วย

Flow Search System

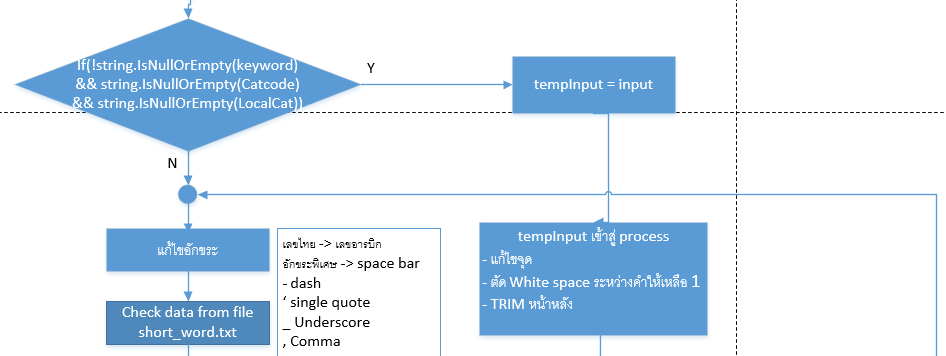


เริ่มต้นการทำงานของระบบ Search มีการ Load file ที่เป็น Text file มา 5 file จะมี synonym.txt, short\_word.txt, ip.txt, token.txt, text\_to\_cat.txt เพื่อมาอ่าน config ที่ได้ Set ค่าใว้

หลังจากนั้นก็จะมีการรับ input เข้ามาจาก User เเละจำมีการ Check input ว่า format ที่ input เข้ามานั้น ถูกต้องหรือไม่ ถ้าไม่ถูกต้อง ระบบก็จะ Return 400 bad request เเละจบการทำงานโดยทันที

ถ้า Check format ผ่านเเล้ว ก็จะมีการ Check IP หรือ Token ของ User ว่าตรงตาม file config หรือไม่ ถ้าไม่ตรงก็จะ Return You don’t have permission เเล้วจบการทำงานของระบบ

ถ้า Check Token หรือ IP ผ่าน เเละถ้าAdmin Code 1-3 ไม่มีค่า จะมา Check Admin ใน keyword ถ้ามี Admin ใน keyword ก็จะถูกต้ดออกจาก keyword เเละจะ input admin code เข้ามา เเละส่งไปทำงานในขั้นตอนต่อไป หรือถ้าไม่มี Admin อยู่ใน keyword อยู่เเล้ว ก็ไม่ต้องตัด Admin ออกจาก keyword เเละไม่ต้องเพิ่ม Admin code เข้ามา เเละไปทำงานในขั้นตอนต่อไปได้เลย



เมื่อระบบมีการ Check Config ครบทุกอันแล้ว ระบบก็จะมาเช็ค Keyword CatCode และ LocalCode จาก Input ที่รับเข้ามา จะ Check ว่าถ้าตัว Keyword ไม่ได้มีค่าว่าง และ CatCode กับ LocalCode นั้นมีค่าว่างจาก Input ที่รับเข้ามา ก็จะถูกนำไปค้นหาที่ไฟล์ text\_to\_cat แต่ก่อนจะไปทำการค้นหาก็จะมีกระบวนการ Process โดยมีการทำให้ Input ถูกเก็บเข้า tempinput ก่อน หลังจากนั้นจะเข้าสู่ Process ในการแก้ไขจุด มีการตัดแต่ง White Space ระหว่างคำ มีการตกแต่ง TRIM หน้าแต่หลังของ Keyword

Diagram

Description automatically generated

เมื่อ Process เสร็จสิ้น tempInput ก็จะถูกส่งมาค้นหาคำจากไฟล์ text\_to\_cat ว่าคำตรงกับในไฟล์ ก็จะเพิ่ม LocalCatCode เข้ามา และจะส่งไปเตรียมจัด Format เพื่อส่งให้โปรแกรม Search engine ต่อไป ถ้าคำไม่ตรงกับไฟล์ text\_to\_cat ก็จะกลับไปใช้ Input ต้นฉบับที่มีแต่ Keyword

Diagram

Description automatically generated

เมื่อทำ text\_to\_cat ไม่ได้และกลับมาใช้ Keyword ต้นฉบับ ก็จะมีการแก้ไขอักขระ เช่น เลขไทยเป็นเลขอารบิก อักษรพิเศษให้เป็น Space bar เมื่อแก้ไขอักขระพิเศษเสร็จ ก็จะนำไป Check ในไฟล์ short\_word ที่เป็นไฟล์คำย่อ

Diagram

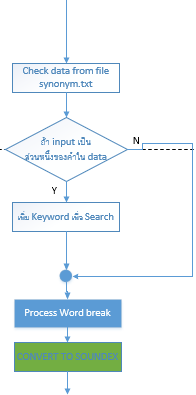
Description automatically generated with low confidence

ถ้าในไฟล์คำย่อไม่เจอ Keyword ก็จะถูกส่งมาทำการ Process ในการแก้ไขจุด การตัด White space ระหว่างคำ การ TRIM หน้าหลัง Keyword และก็จะถูกไป Check ถ้าใน Keyword มีไฟล์ในคำใหญ่

Diagram

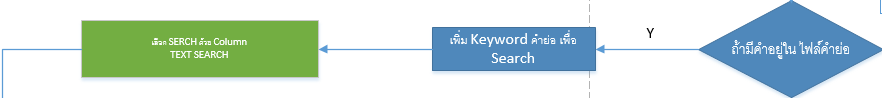
Description automatically generated

เมื่อ Check Keyword แล้วมีไฟล์ในคำใหญ่ ก็จะมาทำการ Search ด้วย Lucene จะมีการ Check data จากไฟล์ synonym ถ้า Input เป็นส่วนหนึ่งของคำใน data ก็จะมีการเพิ่ม Keyword เพื่อไปทำการ Search ถ้าไม่ก็จะไม่เพิ่ม Keyword และถูกนำไปจัด Format เตรียมส่งให้กับตัวระบบ Search engine ต่อไป



แต่เมื่อ Check Keyword แล้วไม่มีไฟล์ในคำใหญ่ ก็จะถูกส่งมา Check data ในไฟล์ synonym ถ้าเจอก็จะถูกเพิ่มตัว Keyword เพื่อไป Search ทันที ถ้าไม่มีเจอก็จะไม่มีการเพิ่ม Keyword จากไฟล์ synonym แล้วเมื่อเสร็จทั้งหมด ก็จะส่งไป Process Word break แล้วทำเป็น Soundex และจัด format เตรียมส่งให้กับ Search engine ทำงานต่อไป

Diagram

Description automatically generated

จากเงื่อนไขข้างต้นถ้ามีคำย่ออยู่ในไฟล์คำย่อ ก็จะเพิ่ม Keyword คำย่อเพื่อ Search โดยจะเป็นการ Search ด้วย Colum TEXT SEARCH และมีการ Check data จาก ไฟล์ synonym ถ้า Input เป็นส่วนหนึ่งของคำใน data ก็จะมีการเพิ่ม Keyword เพื่อไปทำการ Search ถ้าไม่ก็จะไม่เพิ่ม Keyword และทำการ Process Word break และถูกนำไปจัด Format เตรียมส่งให้กับตัวระบบ Search engine ต่อไป

**การทำงานฝั่ง Search Engine**

Graphical user interface, application

Description automatically generated with medium confidence

A screenshot of a computer

Description automatically generated with medium confidence

ถ้าได้ Format สำหรับการ Search จาก Web service มาแล้ว ตัวระบบของ Search Engine ก็จะมีการ Check ตัว Catcode ถ้ามี Catcode ส่งมาด้วย ก็จะมีการ Add Filter Catcode เข้าไปเพื่อให้ค้นหาได้แม่นยำยิ่งขึ้น จากนั้นก็จะมี Check Admin ถ้ามี Admin เข้ามาใน Format ก็จะนำไป Query admin code จาก Database ออกมา และเพิ่มเข้าไปใน Filter เพื่อให้ค้นหาได้แม่นยำยิ่งขึ้นเช่นเดียวกัน จากนั้นก็จะ Check Lat,Lon และ Redius ถ้ามีใน Format ก็จะถูกเพิ่มเข้าไปใน Filter เช่นกัน แต่จะเป็น Filter Nearby และทั้งหมดก็จะถูกส่งไปกระบวนการต่อไป

Diagram

Description automatically generated

เมื่อผ่านกระบวนการ Add filter มาแล้ว ก็จะเข้าสู่กระบวนการเลือก Column สำหรับใช้ในการค้นหาโดยมีเงื่อนไขดังนี้

* ถ้า Keyword เป็นภาษาไทยทั้งหมด ก็จะเลือกใช้ Column NTH
* ถ้า Keyword เป็นภาษาอังกฤษทั้งหมด ก็จะเลือกใช้ Column NEN
* ถ้า Keyword ขึ้นต้นด้วยภาษาไทย ก็จะเลือกใช้ Column NTH\_NEN
* หาก Keyword ไม่ตรงเงื่อนไขข้อใดเลย จะเลือกใช้ Column NEN\_NTH

เมื่อผ่านกระบวนการเลือก Column มาแล้วจะทำการค้นหาข้อมูลตาม Column ที่เลือก

**การจัดกลุ่มข้อมูลสำหรับส่งผลลัพธ์กลับไปยัง Web service**

Diagram

Description automatically generated

หลังจากที่ได้ผลลัพธ์จากกระบวนการเลือก Colum แล้ว ตัวผลลัพธ์ที่ได้ออกมานั้นก็จะเป็น Keyword อีกทีสำหรับการนำไปจัดลำดับเพื่อส่งผลลัพธ์ออกไปให้กับผู้ใช้

กระบวนการทำงานก็จะมีการนำ Keyword มาแบ่งออกด้วยเครื่องหมาย | (pipe) จากนั้นจะทำการใส่หมายเลข Group 4 ไว้ที่ทั้ง 2 Column คือ Goup all , Group name และก็ใส่เลข Group 2 ใน Colum text\_search sortname จากนั้นจะนำ Keyword ที่แบ่งแล้วมาเทียบกับข้อมูลในตาราง

**ความหมายของเลขกลุ่ม**

|  |  |
| --- | --- |
| หมายเลขกลุ่ม | ความหมาย |
| 0 | token ครบและเรียงโดยที่ไม่มีอะไรมาแทรกกลองระหว่างกัน |
| 1 | token ครบ และเรียงแต่มีคำแทรกได้ |
| 2 | token ครบไม่เรียง |
| 3 | token ไม่ครบ |
| 4 | ค่าเริ่มต้น |

**ในกรณีที่ Token ครบ**

1. Case 1 หาก token ตรงทุกตัวใน Column Display name และเรียงจะใส่เลขกลุ่มเป็น 0
2. Case 2 หาก token ครบแต่ตำแหน่งเรียงแต่มีคำแทรก จะใส่เลขกลุ่มเป็น 1
3. Case 3 หาก token ครบแต่ตำแหน่งไม่เรียง มีคำแทรก จะใส่เลขกลุ่มเป็น 2

**ในกรณีที่ Token ไม่ครบ**

หากไม่ได้เข้า case ไหนเลยก็จะใส่เลขกลุ่มเป็น 3