**Documentation**

Diagram

Description automatically generated

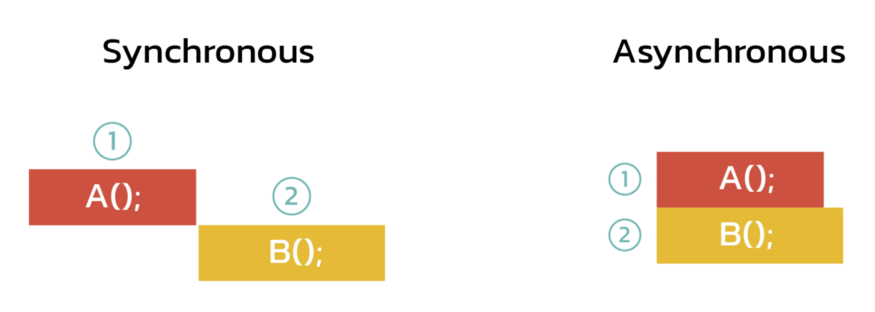
**Flow diagram กระบวนการทำงานของโปรแกรม**

Diagram, schematic

Description automatically generated

**ทำไมถึงเลือกเขียนโค้ดใหม่?**

1. โค้ดของเดิมถูกเขียนด้วย Python 2.7 ทำให้ต้องมีการแก้ไขบาง Syntax เพื่อให้สามารถใช้ได้กับ Python 3 ได้ เนื่องจาก Python 3 เป็นเวอร์ชั่นที่เป็นปัจจุบัน ใช้งานได้กับหลากหลาย Field มี Library ที่รองรับมากมาย และในอนาคต Python 2 อาจจะเริ่มไม่ค่อยมีใครนิยมใช้แล้วก็ได้
2. การอ่านไฟล์ใช้ได้แค่กับไฟล์สกุล .txt ได้เพียงอย่างเดียว เนื่องจากไม่พบการเรียกไฟล์ประเภทอื่น ๆ
3. กระบวนการ Preprocessor และ Tokenizer เหมือนยังทำไม่เสร็จ คลัง Stop word ที่ใช้เก็บลักษณะไฟล์เก่าไม่ได้มาจากการดึงมาจาก module ที่สามารถอัพเดทใหม่ในภายหลังได้
4. ในโค้ดตัวใหม่มีการนำแพคเกจ concurrent.futures ที่มีการทำงานแบบ Asynchronous หรือการทำงานแบบที่รันชุดคำสั่งถัดไปทันทีโดยไม่ต้องรอการทำงานของชุดคำสั่งก่อนหน้า เพื่อให้การทำงานเป็นไปอย่างรวดเร็วมากขึ้น



[Synchronous กับ Asynchronous ต่างกันอย่างไร | by Supanut Dokmaithong | Devbake | Medium](https://medium.com/devbake/synchronous-%E0%B8%81%E0%B8%B1%E0%B8%9A-asynchronous-%E0%B8%95%E0%B9%88%E0%B8%B2%E0%B8%87%E0%B8%81%E0%B8%B1%E0%B8%99%E0%B8%AD%E0%B8%A2%E0%B9%88%E0%B8%B2%E0%B8%87%E0%B9%84%E0%B8%A3-6e6d42b94e4f)

1. โครงสร้างโฟลเดอร์ที่ใช้จัดเก็บไฟล์ยังไม่ค่อยเป็นระเบียบ และยังมีไฟล์ที่ดูซ้ำซ้อนไม่น่าจะได้ถูกนำไปใช้งานจริงยังคงเก็บไว้อยู่

**ข้อจำกัด**

1. โครงสร้างข้อมูลที่ใช้ import เข้าโปรแกรม จะต้องมีลักษณะดังนี้

Calendar

Description automatically generated

* 1. หัวคอลัมน์ที่ใช้ระบุเป็น Attributes ต้องเป็น 1 แถว เท่านั้น ประกอบด้วยเนื้อหา (Context) และ ป้ายกำกับ (Tag) เป็นอย่างน้อย 2 คอลัมน์ และเพื่อให้การทำงานสะดวกขึ้น ควรตั้งชื่อคอลัมน์ข้อมูลที่เป็นข้อความส่วนที่จะใช้ train ว่า ‘context’ และ label ที่จะใช้ระบุผลลัพธ์เป็น ‘tag’ ซึ่งสามารถแก้ไขข้อตกลงเรื่องชื่อคอลัมน์ในภายหลังได้
  2. ในขณะนี้สามารถนำเข้าข้อมูลที่จะใช้เทรนได้แค่เพียงไฟล์ประเภทที่มีการรูปแบบการ separate ที่ชัดเจนเท่านั้น เช่น excel, csv, tsv และ txt เป็นต้น **ซึ่งถ้าจะให้ดีที่สุดควรเป็นไฟล์ประเภท tsv**

1. อัลกอริทึมที่เลือกมาใช้ในการ train ประกอบไปด้วย
   1. Decision Tree
   2. Random Forest
   3. Support Vector Machine
   4. Stochastic Gradient Descent Classifier
   5. Multinomial Naïve Bayes
   6. Logistic Regression
   7. Multilayer Perceptron
   8. Nearest Centrioid
   9. AdaBoost Classifier

อัลกอริทึมนอกเหนือจากนี้จะไม่ได้ถูกนำมาใช้ในการประมวลผลด้วย

1. หกด