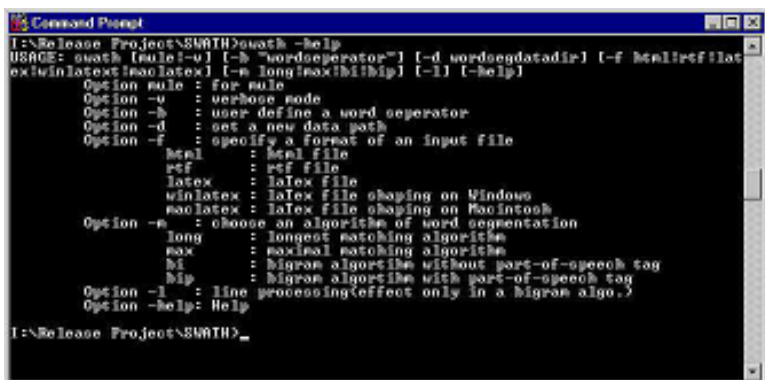


ซอฟต์แวร์ตัดคำและกำกับหน้าที่ คำภาษาไทยอัตโนมัติ



```
I:\Release Project\SWATH>swath -help
USAGE: swath [mode [-u] [-b "wordseparator"]] [-d wordengdatadir] [-f Metafile]
ext\winlatex\maclatex] [-m long|max|hi|big] [-l] [-help]
Option -m : for mode
Option -u : verbose mode
Option -b : user define a word separator
Option -d : set a new data path
Option -f : specify a format of an input file
      Meta : Meta file
      rtf : rtf file
      latex : LaTeX file
      winlatex : LaTeX file chopping on Windows
      maclatex : LaTeX file chopping on Macintosh
Option -m : choose an algorithm of word segmentation
      long : longest matching algorithm
      max : maximal matching algorithm
      hi : bigram algorithm without part-of-speech tag
      big : bigram algorithm with part-of-speech tag
Option -l : line processing effect only in a bigram algo.
Option -help: Help

I:\Release Project\SWATH>
```

ระบบตัดคำภาษาไทยและกำกับหน้าที่คำอัตโนมัติ โดยใช้
โมเดลทางสถิติซึ่งเป็นคลังข้อความที่ทำการตัดคำและกำกับ
หน้าที่คำโดยนักภาษาศาสตร์

คุณสมบัติ

1. เป็นระบบตัดคำภาษาไทยและกำกับหน้าที่คำอัตโนมัติโดยใช้
โมเดลทางสถิติ เทคนิคที่ใช้ได้แก่ ฮิดเดนมาร์คอฟโมเดล (Hidden
Markov Model) และวิเทอร์บีอัลกอริทึม (Viterbi Algorithm)
สำหรับคลังข้อความที่นำมาใช้ในการหาค่าสถิติต่างๆ คือคลัง
ข้อความอรรถคิด ซึ่งเป็นคลังข้อความที่ทำการตัดคำและกำกับหน้าที่
คำโดยนักภาษาศาสตร์
2. เป็นผลงานที่อยู่ในรูปแบบต่างๆ ซึ่งนำไปใช้งานได้หลายแบบ
เช่น เป็นโปรแกรมที่ทำงานบนดอส บนไมโครซอฟท์วินโดวส์ 95 ขึ้นไป
บนระบบปฏิบัติการตระกูลยูนิกซ์ ลินุกซ์ ไลบรารีภาษาซี และไดนา-
มิกลิงค์ไลบรารี (Dynamic Link Library)

3. การตัดคำและกำกับหน้าที่คำอัตโนมัติสามารถนำไปประยุกต์เพื่อพัฒนาเทคโนโลยีต่างๆ ที่เกี่ยวกับการประมวลผลภาษาไทย เช่น การนำไปใช้ร่วมกับเท็กซ์อีดิเตอร์ (Text Editor) เว็บเบราว์เซอร์ (Web Browser) ระบบอ่านเสียงจากข้อความ (Text to speech) หรือระบบแปลภาษาไทยเป็นภาษาอังกฤษ เป็นต้น

ประโยชน์ที่ได้รับ

เป็นพื้นฐานในการพัฒนาเทคโนโลยีของซอฟต์แวร์ประมวลผลภาษาไทย ทำให้ได้ซอฟต์แวร์ที่มีคุณสมบัติดี

วิจัยและพัฒนาโดย: งานวิจัยและพัฒนาสาขาสารสนเทศซอฟต์แวร์ ภาษาและวิศวกรรมความรู้
โครงการพิเศษและพัฒนาการตัดคำภาษาไทยและกำกับหน้าที่คำอัตโนมัติ
สอบถามข้อมูลทางเทคนิคเพิ่มเติมได้ที่หน้า 249

รหัสผลงาน: BI102-43

ติดต่อสอบถามได้ที่: ฝ่ายพัฒนาธุรกิจและอุตสาหกรรม ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และ
คอมพิวเตอร์แห่งชาติ

โทร. (662) 644-8150..99 ต่อ 610, 656..9

โทรสาร (662) 644-8122, 644-8137..8