

การระบุความผิดของผู้ต้องหาจากการรายงานการสอบสวน โดยใช้การประมวลผลภาษาธรรมชาติ

Legal texts analysis for identifying suspect's violated law sections

By using Natural Language Processing technique

ที่มาและความสำคัญ

ในปัจจุบันขั้นตอนการทำการสำรวจของเจ้าหน้าที่ต้องคือ บันทึกคำให้การจากพยาน เพื่อนำคำให้การนั้น ไประบุว่าผู้ต้องหาทำผิดมาตราใดบ้าง เพื่อยืนยันฟ้องต่อศาล โดยขั้นตอนข้างต้นทำโดยประสบการณ์และความสามารถของเจ้าหน้าที่ทั้งหมด ซึ่งขั้นตอนที่กล่าวมาใช้เวลาค่อนข้างมาก และอาจจะเกิดข้อผิดพลาดได้เนื่องจากในปัจจุบันเทคโนโลยีเกี่ยวกับการประมวลผลด้านภาษา มีความก้าวหน้าอย่างมาก ด้วยเหตุนี้หากเรานำเทคนิคการประมวลผลภาษาธรรมชาติมาช่วยระบุความผิด จะทำให้ช่วยเพิ่มความครบถ้วน ช่วยลดภาระและเวลาในการทำงานของเจ้าหน้าที่ได้

เทคนิคที่ใช้

การให้ค่าน้ำหนักคำ เป็นการวัดค่าน้ำหนักของคำที่เอาไว้ประเมินความสำคัญของคำนั้นๆ โดยคำที่เป็นตัวแทนเอกสารที่ดีมักจะปรากฏในเนื้อหาของเอกสารเฉพาะฉบับนั้น และปรากฏอยู่ในชุดของเอกสารอื่น วิธีที่นิยมใช้คือ TF-IDF (Term Frequency - Inverted Document Frequency)

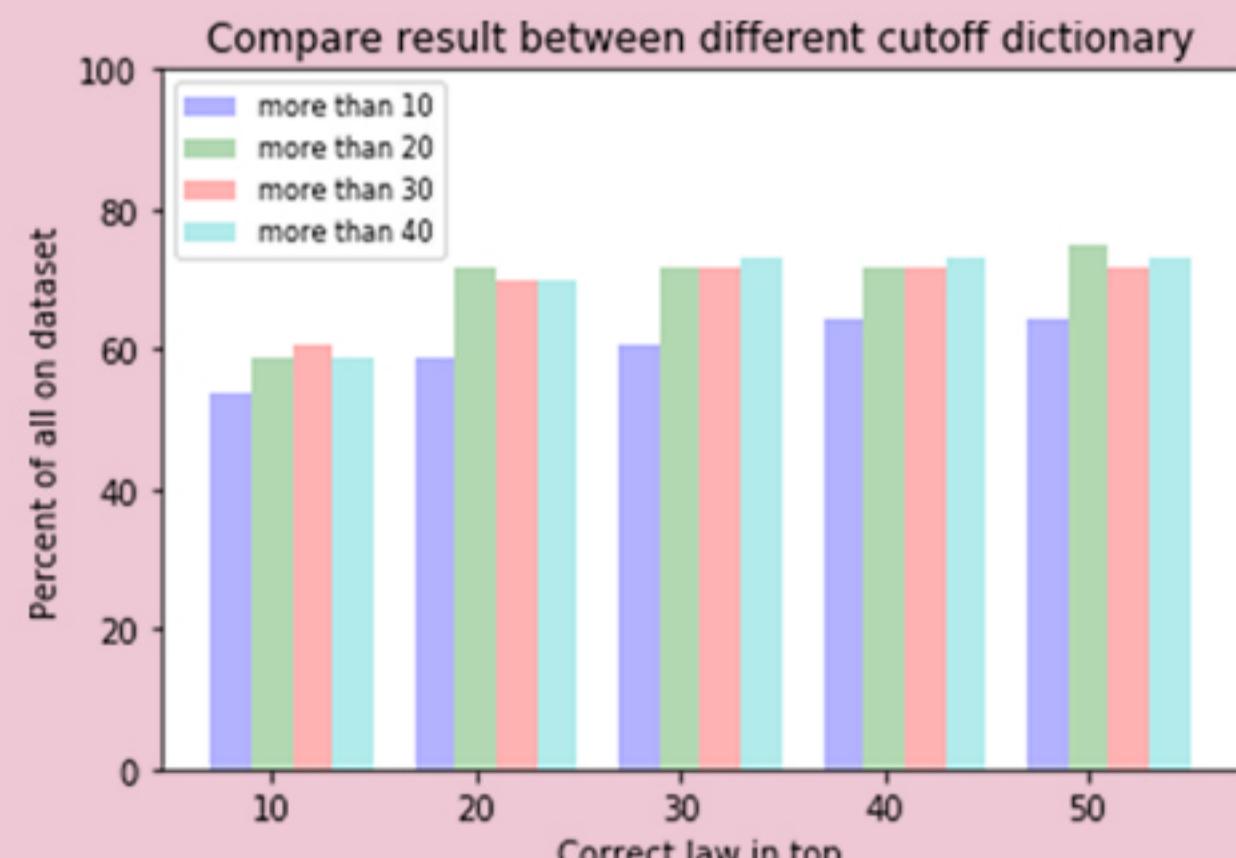
$$tf(term, document) = \frac{f(term, document)}{\sum_{term' \in document} f(term', document)}$$

IDF คือส่วนที่แสดงว่า ถ้าคำนี้ปรากฏในหลายเอกสาร ความสำคัญจะลดลง

$$idf(term, allDocuments) = \log \frac{N}{df(t)}$$

លោកស្រី

จากการทดลองใช้ Dictionary จากคำการที่ตัดคำที่เกิดมากกว่า 10%, 20%, 30% และ 40% จะเห็นว่ามีมาตราที่ถูกอยู่ในอันดับ 10 - 50 มาตราแรก ดังนี้



ໂຄຮງສ້າງຮະບບ

ประมวลกฎหมายอาญา

มาตรา 276 - 366

คำให้การของพยานจาก สำนวนของเจ้าหน้าที่

สร้าง Dictionary จากคำที่พบใน กฏหมายน้อยกว่า 10%, 20%, 30% และ 40%

สร้าง vector ของ TF - IDF ของแต่ละมาตราตาม Dictionary

สร้าง vector ของ TF - IDF จากคำในสำนวนตาม Dictionary

หาค่า cosine similarity ของ vector เพื่อให้คะแนนความเหมือน

แสดงมาตราต่างๆ ที่น่าจะเป็นความผิด
หลังจากคำนวณเรียงตามค่าความเหลื่อมล้ำ

สภาพตัวอย่างของระบบ

Legal texts analysis for identifying suspect's violated law sections

By using Natural Language Processing technique

Please enter text you want to analyze

สอบสวน ผู้ก่อการฯ นายชาครุชช์ วังแพร์ว่า ผู้เสียหาย ให้เพื่อความเรื่องที่เกิดขึ้น พบร้านและบ้านว่าเมื่อรับที่ 19 กุมภาพันธ์ 2560 เวลาประมาณ 02.00 น. ได้ขอร้องว่าคนนั้นที่ก่อเรื่องดังต่อไปนี้ คุณเดวิน มีผละ หมายเหตุจะเบิกเงิน กอน751 พินิตไอล หมายเหตุด่วน อี 100 ห้องเลขที่00036476 หมายเหตุเงื่อนไข อี100 ห้อง 00036476 จำนวน 1 ดัน ราคา 40,000 บาทให้บริการเดินทางร้านซึ่งจะ แล้วรับไป หนาหนักด้วยเวลาประมาณ 08.00น. ผู้เสียหายกลับลงมาที่จุดรถกับบ้านว่าขอ จัดอาหารด้วยสูญเสียหายให้คิดอย่าง จึงมาถือทุกกรณีพนักงานและลูก สอบถามด้วยตัวเอง บ้าว่า พอกานได้รับค่าใช้จ่ายแล้วก็ตาม หารดูด้วยตัวเองว่ามีอยู่หรือเปล่า ถูกประชุมว่า และผู้ก่อการฯความผิดมาดำเนินการตามกฎหมาย อีเช่นกันได้ ทำการสืบสวนมาโดยตลอดและการเป็นเวลานี้ คือความผิดด้วยการที่มีผลลัพธ์ในกระบวนการว่าถูกนำไปเป็นผู้ก่อการฯความผิด และตรวจสอบด้วย ต้องล่าสุดไปอยู่ที่ใด หากพยานทราบว่าถูกนำไปเป็นผู้ก่อการฯความผิดและทักษิณต้องการจะไปอยู่ที่ใด จะแจ้งให้พนักงานและบ้านทราบทันที บัญชีห้ามดูถูกประชุมว่าจะ รายงานการของตนที่จะดำเนินการ

Code	Description	SimScore
335(1)	ผู้ไม่มีอาชญากรรม ในทางกฎหมาย	2.23
343	ดำเนินการค่าความเสียหาย มาตรา 341 ให้กับเจ้า ดำเนินการและต้องการจะคืนเงิน ที่ได้รับไปจาก พนักงาน หรือเจ้าของบ้าน ความที่ ถือความผิดให้เจ็บ ประชุม ถูกทำให้เจ็บ ไม่เกินที่เจ็บ หรือปรับไม่เกินหนึ่งหมื่น บาท หรือให้เจ้าที่ปรับให้เจ็บ สำหรับมีผลต่อการดำเนินการ ต่อ ค่าเสียหาย ต้องค่าใช้จ่าย มาตรา 342 อนุญาต หนึ่งหมื่นบาทเท่านั้น ถูกต้อง ดำเนินการให้เจ็บก็ต้องคืนเงินเท่านั้น เช่นเดียวกันที่จะหักค่าใช้จ่าย หนึ่งหมื่นบาทเท่านั้น	0.37

ผู้ไม่มีอาชญากรรม ในทางกฎหมาย

ANALYZE

10 rows ▾ | C < 1-10 of 34 > |>



ผู้จัดทำโครงการ : นางสาวมนาธิยา ธรรมวงศ์ รหัสประจำตัว 5810504531

อาจารย์ที่ปรึกษา : อ.ดร. หัชทัย ชาญแล้ว

ห้องปฏิบัติการ : หน่วยปฏิบัติการวิจัยเชี่ยวชาญเฉพาะการประมวลผลภาษาธรรมชาติ และเทคโนโลยีสารสนเทศอัจฉริยะ

