

หมวด 31

โปรแกรมตรวจสอบการคัดลอกเอกสารภาษาไทย (Thai Plagiarism Detection)

หลักการและเหตุผล

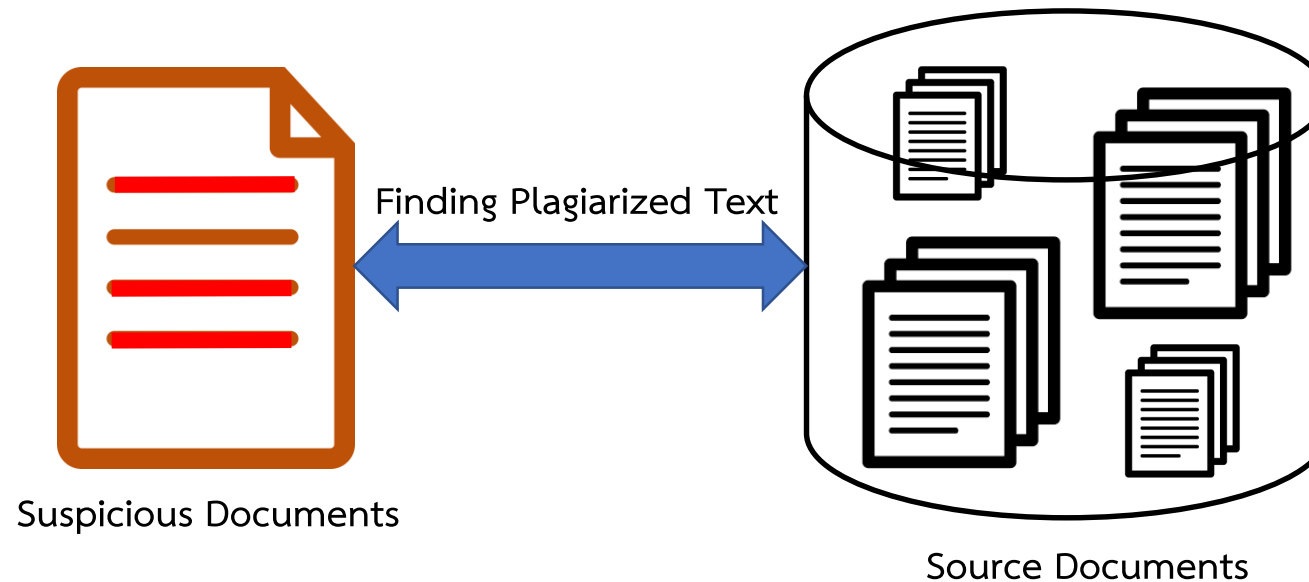
- เอกสารจำนวนมากจะอยู่ในรูปแบบของเอกสารอิเล็กทรอนิกส์และเผยแพร่แบบออนไลน์ ทำให้ง่ายและสะดวกในการคัดลอกข้อความทั้งหมดหรือบางส่วนของเอกสาร
- การตรวจสอบการคัดลอกด้วยมือ (Manual Approach) ทำได้ยาก ต้องอาศัยเวลาและประสบการณ์ของผู้ตรวจ

วัตถุประสงค์ของการแข่งขัน

- เพื่อส่งเสริมความคิดและจินตนาการที่สร้างสรรค์ในการพัฒนาระบบตรวจสอบการคัดลอกเอกสารอัตโนมัติ
- เพื่อสร้างและพัฒนาบุคลากรที่มีความรู้ความสามารถทางด้านการประมวลผลภาษาไทย โดยเฉพาะการพัฒนาระบบตรวจสอบการคัดลอกเอกสารอัตโนมัติ
- เพื่อให้ได้เครื่องมือสำหรับตรวจสอบการคัดลอกเอกสารอัตโนมัติที่มีความหลากหลาย ตลอดจนสามารถนำไปใช้งานได้จริง
- เพื่อแบ่งปันความรู้อันจะนำไปสู่การต่อยอดงานวิจัยขั้นสูงต่อไป

หัวข้อการแข่งขัน

- ให้ข้อมูลที่ประกอบด้วย 1) ชุดเอกสารต้องสงสัย (Suspicious Documents) และ 2) ชุดเอกสารต้นฉบับ (Source Documents)
- ผู้เข้าแข่งขันพัฒนาเครื่องมือเพื่อหาข้อความและแหล่งข้อมูลที่มีการคัดลอก



ตัวอย่างข้อมูล

- คลังข้อมูลประกอบด้วย 1) ชุดเอกสารต้องสงสัย (Suspicious Documents) และ 2) ชุดเอกสารต้นฉบับ (Source Documents)
- เอกสารต้องสงสัยอาจประกอบด้วยข้อความที่มีการคัดลอกจากแหล่งเดียวหรือหลายแหล่งก็ได้
- ข้อความที่มีการคัดลอกอาจมีการคัดลอกมาจากหลายวิธี ได้แก่
 - การลบคำ (Deletion)
 - การแทรก (Insertion)
 - การแทนที่ (Replacing)

รูปแบบการคัดลอก

- ลอก/ตัดแปะ (Copy and Paste) NSC 2018
- เปลี่ยนรูปภาษาที่คัดลอกมา (Language Changes) มี 4 ระดับ
 - ระดับหน่วยคำ (Morpheme) เช่น นักเขียน (2 หน่วยคำ -> นัก+เขียน)
 - ระดับคำ/วลี (Lexicon)
 - ระดับโครงสร้าง (Structure)
 - ระดับความหมาย (Meaning/Semantics)

ตัวอย่างข้อมูล (ต่อ)

- ระดับคำ/วลี แทนที่ด้วยคำพ้องความหมาย (syn), ลบคำ (del), เพิ่มคำ (ins)
S_{src}: การมีที่พักอาศัยเป็นความต้องการขั้นพื้นฐานอย่างหนึ่งของการอยู่รอดของมนุษย์
S_{pla}: การมีที่อยู่อาศัย_{syn}นับ_{ins}เป็นความต้องการขั้นแรก_{syn}อย่างหนึ่ง_{del}ของการอยู่รอดของมนุษย์
S_{result}: การมีที่อยู่อาศัยนับเป็นความต้องการขั้นแรกของการอยู่รอดของมนุษย์

ขั้นตอนการสมัคร



ขอบคุณครับ

ติดตามข่าวสารและดูรายละเอียดเพิ่มเติมได้ที่

<http://www.copycatch.in.th>