

## Code Clone Detector

หัวข้อนี้ใช้โมเดล code2vec ที่ได้รับการฝึกฝนมาแล้วเพื่อแปลงโค้ด snippet เป็นตัวแทนเชิงตัวเลข (code vector) จากนั้นจึงใช้ค่าสัมประสิทธิ์ความคล้ายคลึงเชิงโคไซน์ (cosine similarity) เพื่อวัดความคล้ายคลึงเชิงความหมายระหว่าง code vectors

ค่าสัมประสิทธิ์ความคล้ายคลึงเชิงโคไซน์เป็นค่าที่อยู่ระหว่าง 0 ถึง 1 ค่าที่ใกล้ 1 แสดงว่า code vectors สองตัวมีความคล้ายคลึงกันมาก ค่าที่ใกล้ 0 แสดงว่า code vectors สองตัวมีความแตกต่างกันมาก

### ใช้หลักการต่อไปนี้ในการทำงาน

- โมเดล code2vec ที่ได้รับการฝึกฝนมาแล้ว (Pretrained Model)
- แปลงโค้ด snippet เป็น code vectors
- คำนวณค่าสัมประสิทธิ์ความคล้ายคลึงเชิงโคไซน์สำหรับคู่ของ code vectors ที่ต้องการ
- ตีความค่าสัมประสิทธิ์ความคล้ายคลึงเชิงโคไซน์เพื่อเข้าใจความสัมพันธ์ระหว่างโค้ด snippet

### โครงการนี้สามารถให้ประโยชน์ต่างๆ แก่โปรแกรมเมอร์และนักวิจัย ตัวอย่างเช่น

- ช่วยให้โปรแกรมเมอร์ค้นหาโค้ด snippet ที่คล้ายคลึงกันได้ง่ายขึ้น
- ช่วยให้นักวิจัยระบุประเภทของโค้ด snippet ที่ใช้บ่อยที่สุด
- ช่วยให้วิศวกรซอฟต์แวร์ตรวจสอบความถูกต้องของโค้ดที่เขียนโดยอัตโนมัติ

### นอกจากนี้ โครงการนี้ยังสามารถนำไปใช้เพื่อวัตถุประสงค์อื่นๆ เช่น

- สอนโปรแกรมเมอร์เกี่ยวกับโครงสร้างและรูปแบบของโค้ด
- พัฒนาเครื่องมือสำหรับการสร้างโค้ดอัตโนมัติ
- ปรับปรุงการบำรุงรักษาซอฟต์แวร์

โดยสรุป โครงการนี้เป็นเครื่องมือที่มีประโยชน์สำหรับโปรแกรมเมอร์และนักวิจัยที่ต้องการพัฒนาซอฟต์แวร์ได้อย่างมีประสิทธิภาพ