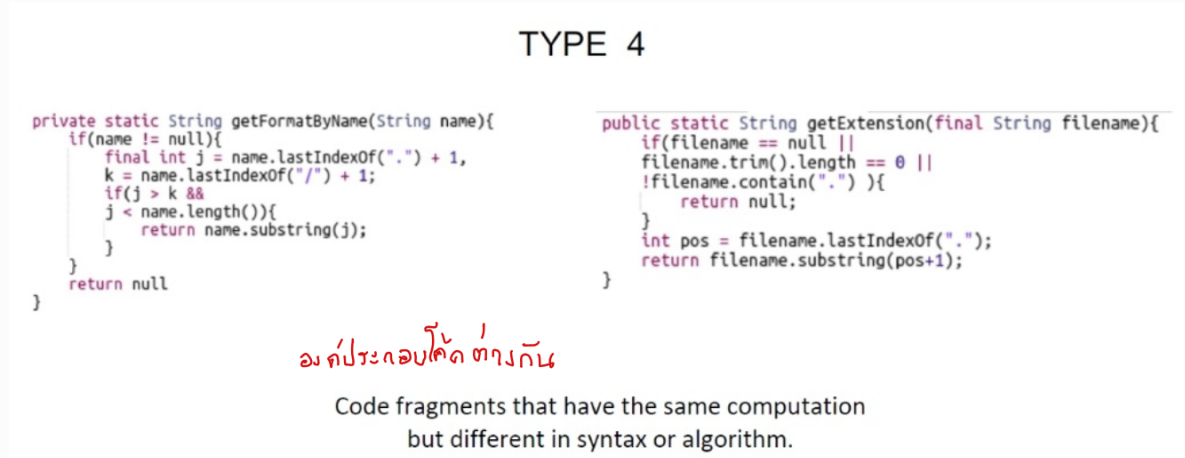
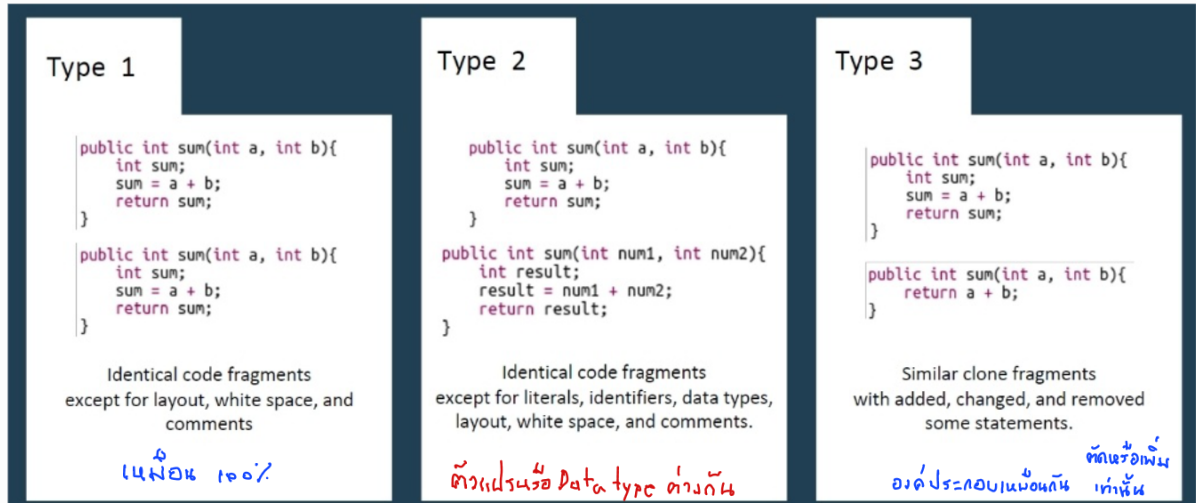


What is Code clone : Two code fragments form a clone pair if they are similar enough according to a given definition of similarity

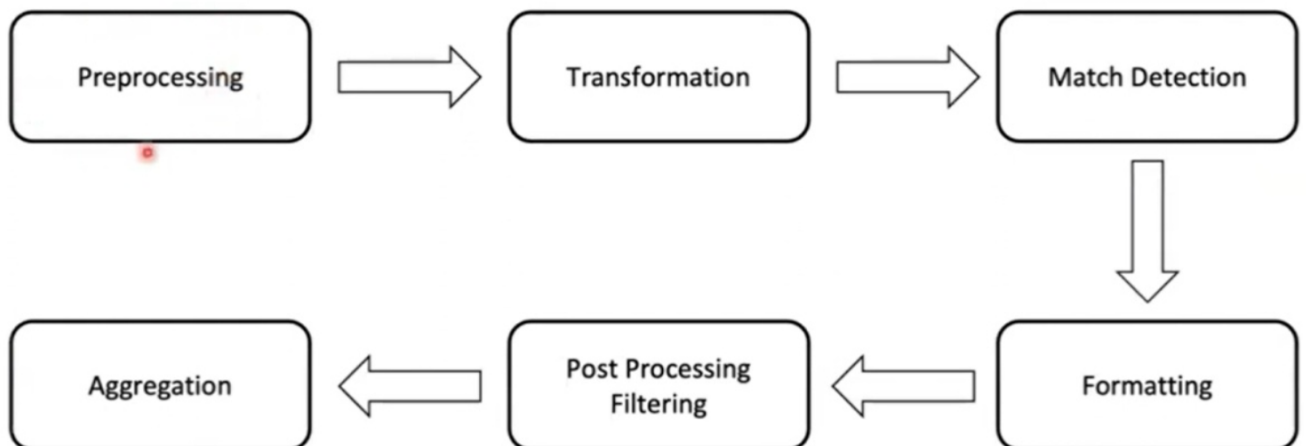
ประเภท :



ทำไมต้องมี Code detector : มีระบบจัดการโค้ด

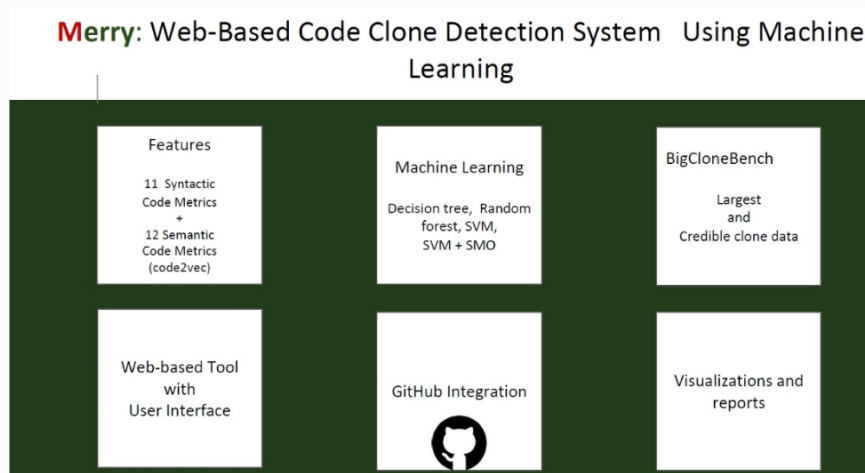
Process :

- สมมุติว่าถ้าได้โค้ดแล้วเรา Copy มา อาจส่งผลต่อภาพในยูเอ
- กันการขโมย Code ลับกับ Code ที่ private ไว้
- Code clone บางตัวก็มีประโยชน์



Model : Merry

- เป็น AI ที่พัฒนา Web-based
- feature คำนวณหาฟังก์ชัน
- แยก type 1, type 3



Data collection

ใช้ BigCloneBench Dataset ซึ่งใช้ Script มาประมวลผล

Training Data [8]				
True Clone Pairs (22,663 pairs)				False Clone Pairs
Type 1	Type 2	Very Strong Type 3	Strongly Type 3	
13,750	3,104	1,207	4,602	22,663

Testing Data				
True Clone Pairs (4,724 pairs - stratified sampling)				False Clone Pairs
Type 1	Type 2	Very Strong Type 3	Strongly Type 3	
2,383	557	307	1,477	18,893

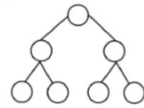
Code metrics Extraction

- Syntactic metrics : ดูโครงสร้างการเขียน
- Semantic metrics : ดูความหมาย ใช้ Code2vec เป็น Feature extractor

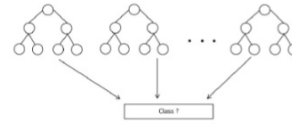
Model

1.3

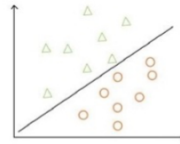
Machine Learning Models



Decision Tree



Random Forest



Support Vector Machine



SVM using SMO

Model Evaluation

Model	Metrics	Precision	Recall	F1-Score
Randomization (baseline)		0.20	0.49	0.28
Decision Tree	Syntactic + Semantic	0.89	0.86	0.87
	Syntactic	0.95	0.72	0.86
	Semantic	0.68	0.87	0.76
Random Forest	Syntactic + Semantic	0.97	0.86	0.91
	Syntactic	0.97	0.80	0.87
	Semantic	0.70	0.87	0.78
SVM	Syntactic + Semantic	0.97	0.85	0.91
	Syntactic	0.97	0.79	0.87
	Semantic	0.62	0.90	0.73
SVM using SMO	Syntactic + Semantic	0.98	0.89	0.93
	Syntactic	0.97	0.69	0.81
	Semantic	0.63	0.90	0.74