AUMENA: 一种基于提示学习的方法命名自动化方法

Jie Zhu, Lingwei Li, Li Yang*, Xiaoxiao Ma, Chun Zuo

The 31st IEEE/ACM International Conference on Program Comprehension (ICPC 2023), 2023

Tel：杨立 yangli2017@iscas.ac.cn 010-62661198

Background

- **Method names** are brief summaries of source code and could indicate the developer's purpose.
- However, **method names** could also be confusing, making programs even harder to understand and more error-prone.

```
public int getBooleanValue()
{
    String value = super.getValue();
    try {
        return Integer.parseInt(value);
    } catch (NumberFormatException e) {
        // TODO: validation handling/logging
        throw (e);
    }
}
```

The original method name is inconsistent with its function and we should replace it with a better name such as 'getIntValue'.
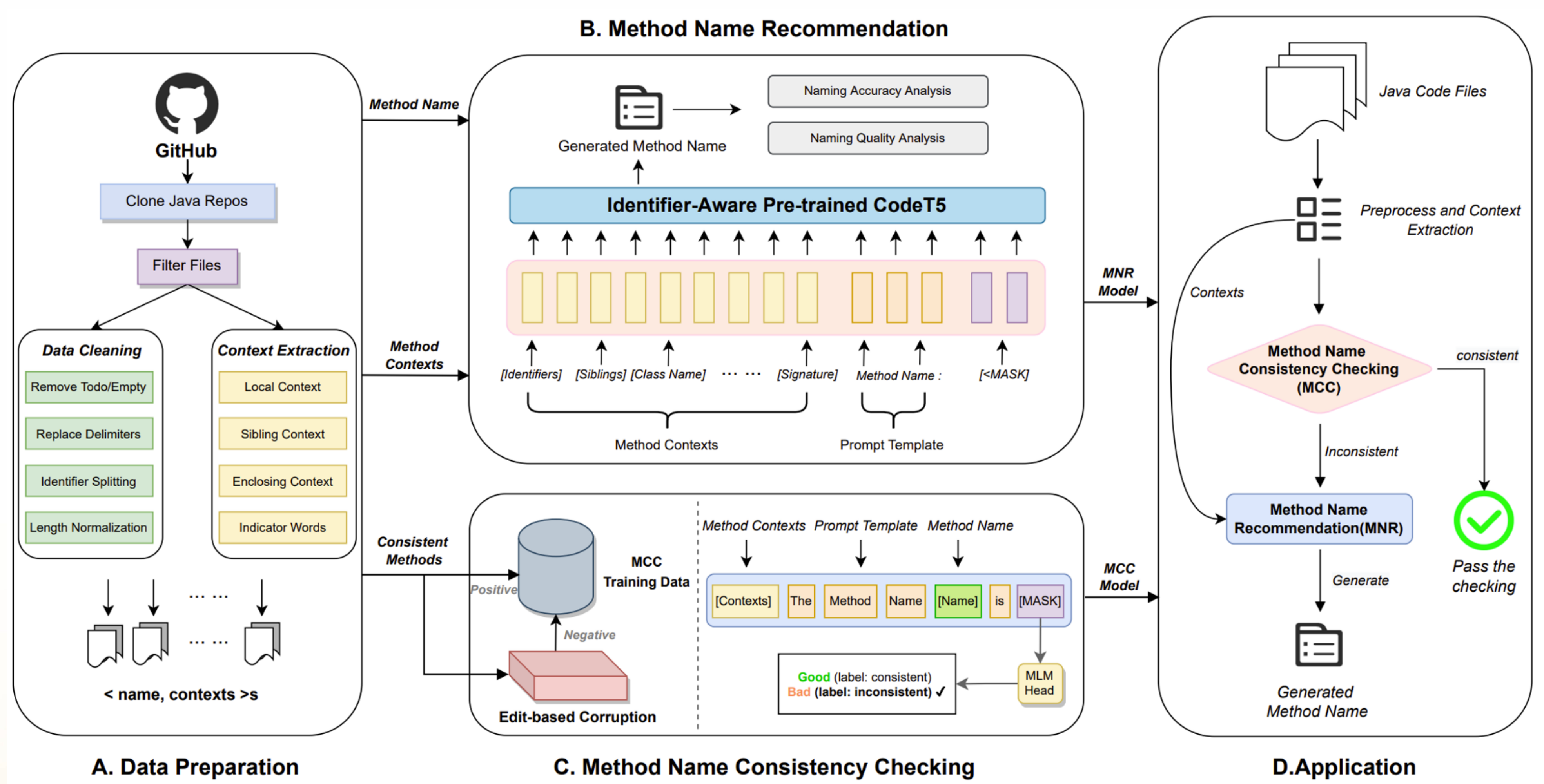
Task1: detect inconsistent method name

Task2: recommend a high-quality method name as candidate

Motivation

- Suboptimal initial understanding.
- Low detection precision.
- Fail to measure semantic consistency.

Approach: AUMENA

- **Data**: 33376k method names from 10k+ Java projects in GitHub.
- **Hard Negative Sampling**: construct inconsistent names from consistent ones.
- **Prompt tuning**: two prompt-based classification and generation models.

Evaluation

- **RQ1: Name Generation Accuracy**
  - AUMENA outperforms all baselines on four datasets when generating method names.

- **RQ2: Inconsistency Detection Accuracy**
  - Compared with previous methods, AUMENA achieves higher precision on inconsistency detection and better overall accuracy.

- **RQ3: Naming Quality**
  - AUMENA could generate method names with similar or even higher quality compared to human-written ones.

Dataset	Approach	Precision	Recall	F1-score	EM Acc
Java-small	Code2vec	23.4	22	21.4	-
	MNire	44.8	38.7	41.5	15.5
	Cognac	67.1	59.7	63.2	-
	AUMENA	69.6	67.6	68.6	44.3
Java-med	Code2vec	36.4	27.9	31.9	-
	MNire	62.0	57.6	59.7	36.2
	Cognac	64.8	57.3	60.8	-
	AUMENA	72.6	71.4	72.0	50.9
Java-large	Code2vec	44.2	38.3	41.6	-
	MNire	63.1	59.0	61	37.4
	Cognac	71.4	61.9	66.3	-
	AUMENA	74.0	73.2	73.6	55.3
MNire's	Code2vec	51.9	39.8	45.1	35.6
	MNire	66.3	62.1	64.2	42.6
	Cognac	70.2	66.8	68.5	-
	GTNM	77.0	74.1	75.6	62.0
	AUMENA	85.1	84.3	84.7	71.0

Legend: Excellent | Good | Fair | Poor

AUMENA	210594 / 84298 / 56622
Human-written	202814 / 87048 / 61399

0   100000   200000   300000