

# PR #22487 完整报告

sgl-project/sclang

[Spec][Ngram] Clean up unused stateless `batchMatch`

合并时间: 2026-04-11 12:52

原文链接: <http://prhub.com.cn/sgl-project/sclang/pull/22487>

## 执行摘要

此 PR 移除了 Ngram 模块中无状态的 `batchMatch` 死代码，清理了 C++、FFI 绑定和 Python 接口中的相关函数，确保代码整洁，无功能影响。

## 功能与动机

在 PR #21243 引入有状态 `batchMatch` 后，残留了无状态的旧代码，可能源于合并冲突解决。为避免代码冗余和潜在混淆，需要清理这些无用代码。PR body 明确表示这是“死代码”，旨在提升代码可维护性。

## 实现拆解

实现基于删除操作，涉及四个关键文件：

- C++ 核心层：`ngram.cpp` 中移除了 `Result Ngram::batchMatch(const std::vector<std::vector<int32_t>>& tokens)` 函数（29 行删除），`ngram.h` 中移除声明。
- FFI 绑定层：`ngram_corpus_ffi.cpp` 中移除 `batch_match` 方法（19 行删除），清理 TVM 绑定。
- Python 接口层：`ngram_corpus.py` 中移除 `match` 方法（17 行删除），保持 API 一致性。

代码变更示例如下（以 `ngram.cpp` 为例）：

```
// 删除的无状态batchMatch实现片段
Result Ngram::batchMatch(const std::vector<std::vector<int32_t>>& tokens) {
    // ... 原逻辑已移除
}
```

## 评论区精华

review 过程简单，reviewer hnyls2002 直接批准，无评论讨论，表明变更被视为低风险清理工作。关联 Issue 评论中显示 CI 测试通过（如 `test_ngram_corpus.py` 和 `test_ngram_speculative_decoding.py`），进一步确认安全性。

## 风险与影响

- 风险：极低。代码确认为死代码，删除无回归风险；性能、安全和兼容性无变化；CI 测试通过提供额外保障。

- 影响：用户无感知，因无状态版本已废弃；系统代码量减少，提升可维护性；团队受益于更简洁的代码库。

## 关联脉络

此 PR 是 PR #21243 的后续清理工作，共同推进 ngram 模块的代码质量。从历史 PR 看，近期有多个涉及 `jit-kernel`、`run-ci` 标签的 PR（如 #22051、#21104），反映该模块持续优化趋势。但本 PR 专注于死代码移除，而非功能增强或性能调整。