2025.05.28.(수)

#### NS3 Code Generator

: LLM 기반의 네트워크 시뮬레이션 코드 자동 생성

202101109 박수화 202102722 최희원

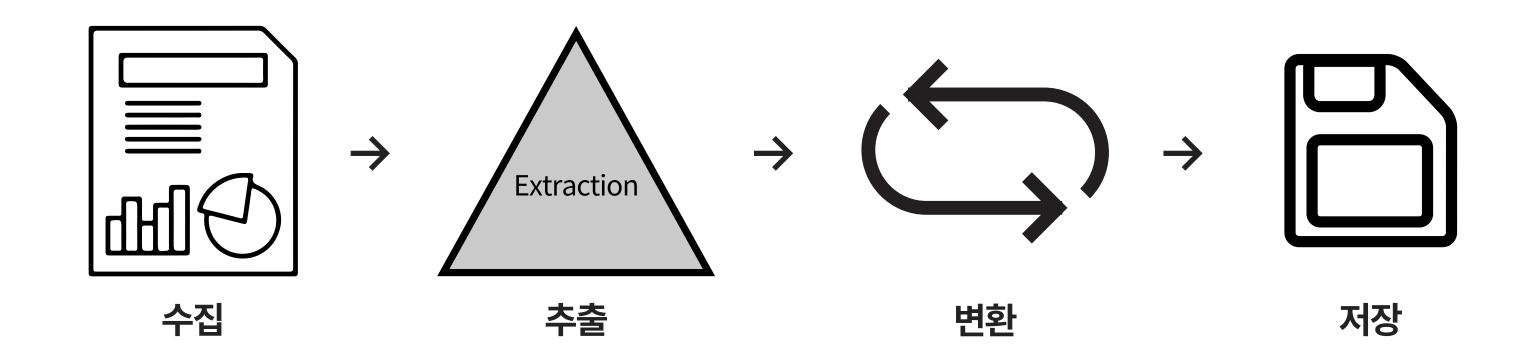
## 목차

1 RAG DB 구축 흐름

Data Crawling (stackoverflow)

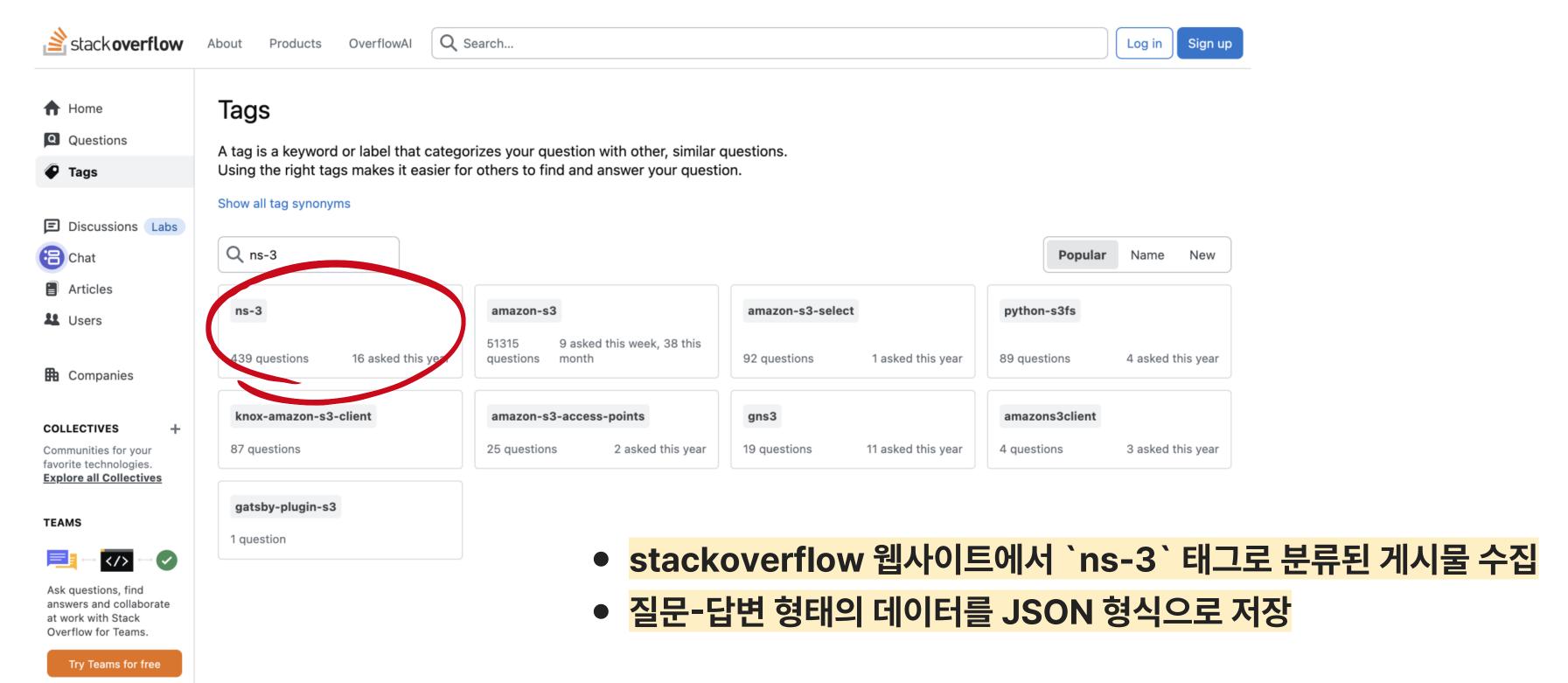
3 Data Crawling (github)

### RAG DB 구축 흐름

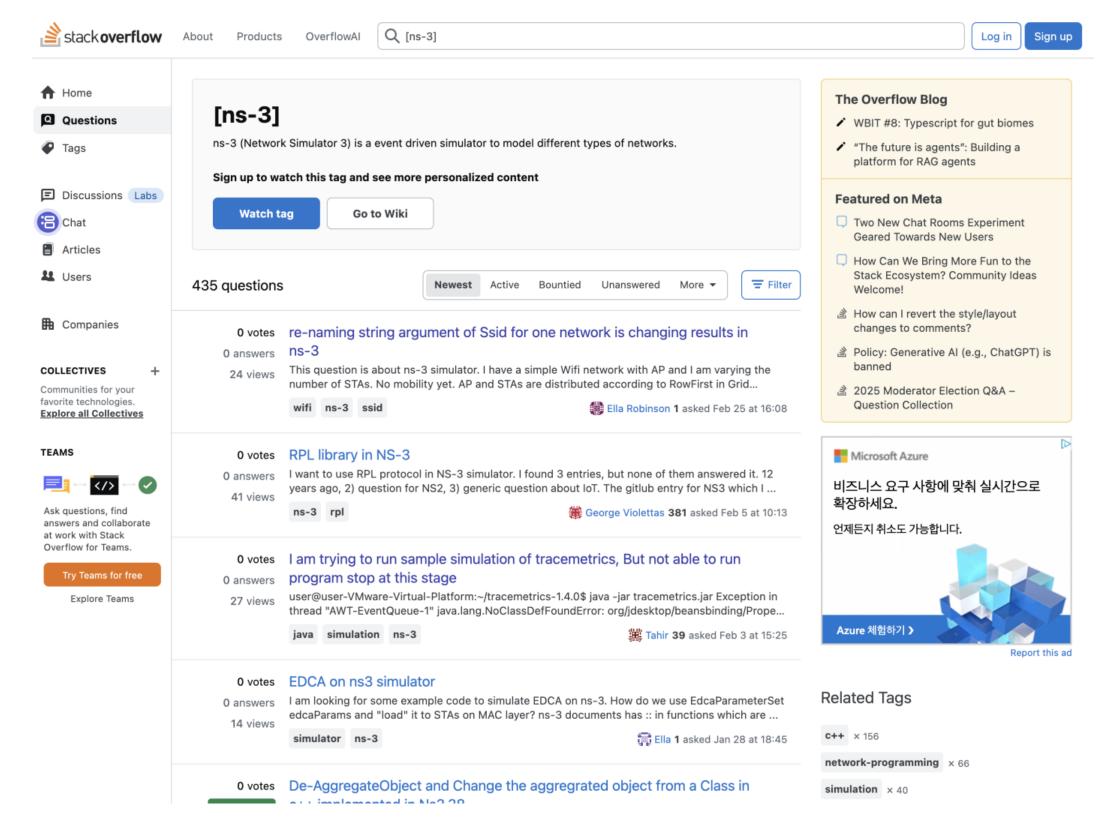


NS-3 태그 기반 Stack Overflow Q&A 데이터 수집

**Explore Teams** 



NS-3 태그 기반 Stack Overflow Q&A 데이터 수집



#### Crawling 방법

- stackoverflow 웹사이트에서 `ns-3` 태그로 분류된 게시물 수집
- 질문-답변 형태의 데이터를 JSON 형식으로 저장
- 질문 페이지에는 여러 개의 답변이 달릴 수 있기 때문에, 질문자가 직접 채택한 답변이 있으면, 답변을 우선적으로 수집
  - → 질문자가 답변을 채택하지 않았을 때는, Stack Overflow에서 사용자들이 투표하여 가장 많은 득표(추천)를 얻
  - → 답변이 없는 경우 제외

은 답변을 선정

#### Crawling 방법

```
import time
import json
import cloudscraper
from bs4 import BeautifulSoup
import re
                                                                                                                                                      stackoverflow.py
# Configuration
BASE_URL
           = 'https://stackoverflow.com'
                                                                                            # Helpers
           = 'ns-3'
TAG
                                                                                            def slugify(title: str) -> str:
                                                                                      33
DELAY
           = 1
                              # Delay between requests (seconds)
                                                                                                """Convert title to URL slug."""
                                                                                      34
MAX PAGES = 30
                              # Number of pages to crawl (15 posts per page)
                                                                                                slug = title.lower()
                                                                                      35
OUTPUT_FILE = 'ns3_qa.json'
                                                                                                slug = re.sub(r"[^a-z0-9\s-]", "", slug)
                                                                                                slug = re.sub(r"[\s-]+", "-", slug).strip('-')
scraper = cloudscraper.create_scraper()
                                                                                      37
HEADERS = {
                                                                                                return slug
    'accept': 'text/html,application/xhtml+xml,application/xml;q=0.9,image/avif,image/webp
                                                                                     39
    'accept-language': 'ko-KR,ko;q=0.9,en-US;q=0.8,en;q=0.7',
                                                                                      40
    'cache-control': 'no-cache',
                                                                                            def get_question_links(page: int) -> list:
                                                                                      41
    'pragma': 'no-cache',
                                                                                                """Fetch question URLs from the tagged listing page."""
                                                                                      42
   'sec-ch-ua': '"Chromium"; v="136", "Google Chrome"; v="136", "Not.A/Brand"; v="99"',
   'sec-ch-ua-mobile': '?0',
                                                                                                url = f"{BASE_URL}/questions/tagged/{TAG}?tab=newest&page={page}"
                                                                                      43
   'sec-ch-ua-platform': '"macOS"',
                                                                                      44
                                                                                                resp = scraper.get(url, headers=HEADERS)
   'sec-fetch-dest': 'document',
                                                                                                resp.raise for status()
                                                                                      45
   'sec-fetch-mode': 'navigate',
                                                                                                soup = BeautifulSoup(resp.text, 'html.parser')
                                                                                      46
    'sec-fetch-site': 'none',
                                                                                      47
   'sec-fetch-user': '?1',
                                                                                      48
                                                                                                links = []
   'upgrade-insecure-requests': '1',
                                                                                                for summary in soup.select('div.s-post-summary'):
   'user-agent': 'Mozilla/5.0 (Macintosh; Intel Mac OS X 10_15_7) AppleWebKit/537.36 (KHT 49
    'cookie': 'prov=17412cb4-95bb-42a0-b540-b9eb92e3817f; __cflb=02DiuFA7zZL3enAQJD3AX8Zzv 50
                                                                                                    a = summary.select_one('h3.s-post-summary--content-title a.s-link')
                                                                                                    if a and a.get('href'):
                                                                                      52
                                                                                                         links.append(BASE_URL + a['href'])
                                                                                      53
                                                                                                return links
```

#### Crawling 방법

stackoverflow.py

```
def parse_question(url: str) -> dict:
         """Parse question page for title, body, and best answer."""
         resp = scraper.get(url, headers=HEADERS)
         resp.raise_for_status()
         soup = BeautifulSoup(resp.text, 'html.parser')
62
         # Title
         title tag = soup.select one('h1 a.guestion-hyperlink')
         title = title_tag.get_text(strip=True) if title_tag else ''
         # Question body
         q_body = soup.select_one('div.s-prose.js-post-body')
         question = q_body.get_text("\n", strip=True) if q_body else ''
70
         # Answer: accepted first, else highest-score
         acc = soup.select_one('div.answer.js-accepted-answer')
             body = acc.select_one('div.s-prose.js-post-body')
             answer = body.get_text("\n", strip=True) if body else ''
         else:
             candidates = []
             for a_div in soup.select('div.answer'):
                 body = a_div.select_one('div.s-prose.js-post-body')
                 text = body.get_text("\n", strip=True) if body else ''
                 score_tag = a_div.select_one('div.js-vote-count')
                     score = int(score_tag.get_text(strip=True))
84
                     score = 0
                 candidates.append((score, text))
             if candidates:
                 answer = max(candidates, key=lambda x: x[0])[1]
         return {'url': url, 'title': title, 'question': question, 'answer': answer}
```

```
# Main
      def main():
          results = []
95
          for page in range(1, MAX_PAGES + 1):
              print(f"[Page {page}] fetching listings...")
97
              try:
                  links = get_question_links(page)
99
              except Exception as e:
100
                  print(f"Failed to fetch page {page}: {e}")
101
                  break
102
              if not links:
103
                  break
104
105
              for link in links:
106
                  print(f" -> parsing {link}")
107
                      qa = parse_question(link)
109
                      # Skip unanswered questions
110
                      if not qa['answer'].strip():
111
                          print(f" skipped {link} (no answer)")
112
                          continue
113
                      results.append(qa)
114
                  except Exception as e:
115
                      print(f" parse error for {link}: {e}")
116
                  time.sleep(DELAY)
117
118
          # Save to JSON
119
          with open(OUTPUT_FILE, 'w', encoding='utf-8') as f:
120
              json.dump(results, f, ensure_ascii=False, indent=2)
121
          print(f"Saved {len(results)} Q&A entries to {OUTPUT_FILE}")
122
123
      if __name__ == '__main__':
          main()
```

#### Crawling 결과

ns3\_qa.json

```
"url": "https://stackoverflow.com/questions/79179369/de-aggregateobject-and-change-the-aggregrated-object-from-a-class-in-c-impleme",
         "title": "De-AggregateObject and Change the aggregrated object from a Class in c++ implemented in Ns3.38",
         "question": "I implemented Vehicle Ad-hoc Network in NS3.38 and used default Node class in NS3. also I created Cluster Object that has some
         "answer": "I think your goal in this question is clustering nodes and getting each cluster from node object. also each cluster has multi not
         "url": "https://stackoverflow.com/questions/79146271/importing-opency-and-tensorflow-lite-into-ns-3-38",
         "title": "Importing OpenCV and TensorFlow Lite into ns-3.38",
10
         "question": "I am currently working on a project that involves using\nOpenCV\nand\nTensorFlow Lite\nwithin the\nns-3.38\nsimulation environm
12
         "answer": "You're missing include headers in scratch/EDEN/node.cc.\n#include \"ns3/network-module.h\""
13
       },
14
15
         "url": "https://stackoverflow.com/questions/79090343/missing-system-dependencies-for-ns3",
         "title": "missing system dependencies for ns3",
16
         "question": "I have been setting up ns-3-allinone on my WSL Python virtual environment using the in-built 'bake' tool. However during the se
         "answer": "Are you running bake from within your virtual environment?\nIf no, then try doing that.\nIf yes, then it could be a limitation o
18
19
20
21
         "url": "https://stackoverflow.com/questions/78645306/ns-3-multiple-wifi-clients-and-one-server-using-udp",
22
         "title": "NS-3 Multiple WiFi clients and one server using UDP",
23
         "question": "I am trying to built a ns-3 simulation with the following characteristics . I want multiple client nodes that will periodically
         "answer": "I don't know if its right but it seems to work I did it with UdpServer\nNS_LOG_INFO(\"Create UdpServer application on AP node.\"
24
25
26
27
         "url": "https://stackoverflow.com/questions/78485567/how-to-make-full-buffer-traffic-in-ns-3-lte-module",
         "title": "How to make full-buffer traffic in NS-3 LTE-module",
28
29
         "question": "I've already read documentation for LTE-module and EPC packets, but I don't understand how can I create full-buffer traffic. I
30
         "answer": "you can use
                                                                                                Time interPacketInterval = MilliSeconds(1); //1ms for
```

# 감사합니다:)