

## ऐसे मामले जहां हमें अभी भी १००००० सर्च बॉक्स की आवश्यकता है

```
jobs:  
  awesome-cv-copy:  
    runs-on: ubuntu-latest  
    steps:  
  
    # ...  
  
    - name: TeX Live 2023  
      if: steps.cache-texlive.outputs.cache-hit != 'true'  
      run: |  
        # TeX Live  
        sudo apt-get update  
        sudo apt-get install -y perl wget xz-utils  
  
    # TeX Live  
    wget http://mirror.ctan.org/systems/texlive/tlnet/install-tl-unx.tar.gz  
    tar -xzf install-tl-unx.tar.gz  
    cd install-tl-*/  
  
    # ...  
  
    - name:     LaTeX  
      run: |  
        sudo /usr/local/texlive/2023/bin/x86_64-linux/tlmgr install etoolbox adjustbox  
  
    - name:  
      run: |  
        kpsewhich etoolbox.sty  
        kpsewhich adjustbox.sty  
  
    - name: Run make awesome-cv-copy  
      run: make awesome-cv-copy
```

मैं ऊपर दिए गए १००००० १०००००० स्क्रिप्ट पर काम कर रहा हूँ।

मुझे १००००० पर etoolbox adjustbox language: YAML के लिए सटीक कोड खोजने की आवश्यकता है।

मुझे निम्नलिखित त्रुटि का सामना करना पड़ा:

2025-01-07T22:34:58.6493408Z

2025-01-07T22:34:58.6493741Z ! LaTeX : adjustbox.sty

2025-01-07T22:34:58.6494172Z

2025-01-07T22:34:58.6494593Z X <RETURN>

2025-01-07T22:34:58.6495322Z ( : sty)

मैं विशेष स्पष्ट से etoolbox adjustbox language: YAML खोज रहा हूँ, और १००००० में परिणाम सीमित हैं, केवल 53 १००० फ़ाइलें हैं जिनमें etoolbox और adjustbox दोनों शामिल हैं। मुझे एक **सटीक मिलान** चाहिए।

हालांकि हम बड़े भाषा मॉडल्स के युग में हैं, फिर भी सटीक मिलान खोजने की आवश्यकता अभी भी महत्वपूर्ण है। यह विशेष रूप से तब सच होता है जब किसी चीज़ का सटीक अर्थ जांचना हो या सटीक कार्यशील कोड ढूँढना हो। इसी तरह, १००००००, १०००००००, या अन्य प्लेटफॉर्म्स भी अर्थ के लिए सटीक खोज पर निर्भर करते हैं। हम १०-जनति परिणाम या छोटी-मोटी गलतियों वाले परिणाम नहीं चाहते हैं।

बड़े भाषा मॉडल को प्रशिक्षित करने के लिए, हम एक ऐसी प्रणाली विकसित कर सकते हैं जो सटीक मिलान ढूँढ़ती है। शायद हम (०००-००००००-००००००) खोज एल्गोरिदम को **ट्रांसफॉर्मर आर्किटेक्चर** के साथ जोड़कर खोज क्षमताओं को बढ़ा सकते हैं। **ट्रांसफॉर्मर** के साथ ००० का उपयोग करने से विशेष कोड खोजों के लिए अधिक सटीक परिणाम प्राप्त करने में मदद मिल सकती है।

वर्तमान में, बड़े भाषा मॉडल ०००० या ०००००० जैसी फ़ाइल भाषाओं के आधार पर फ़िल्टर नहीं कर सकते हैं। हालांकि, वास्तविक दुनिया में जानकारी का एक बड़ा हिस्सा इसी तरह से व्यवस्थित होता है। इसका मतलब है कि हम बड़े भाषा मॉडल को फ़ाइलों का उपयोग करके प्रशिक्षित कर सकते हैं। यदि हम सभी टेक्स्ट डेटा को फ़ाइल प्रकारों के आधार पर व्यवस्थित करते हैं, तो हम मॉडल को उन्हें बेहतर ढंग से समझने के लिए प्रशिक्षित कर सकते हैं। इसलिए, बड़े भाषा मॉडल के लिए, हमें शुरुआत में फ़ाइल भाषाओं को पूर्वनिर्धारित करने की आवश्यकता होगी। डिफॉल्ट रूप से, यह “टेक्स्ट” हो सकता है, लेकिन हम अन्य भाषाओं को भी परिभाषित कर सकते हैं, जैसे कि ०००००० ०००००० करता है। परिणाम फ़ाइलों को लौटाएगा, जैसे कि ०००००० खोज परिणाम करते हैं।

महत्वपूर्ण हिस्सा **फ़ाइल फॉर्मेट** या **एक्सटेंशन** है, न कि फ़ाइल का नाम। यहां कुछ उदाहरण दिए गए हैं:

हालांकि, जब किसी उपयोगकर्ता का प्रॉम्प्ट सामान्य टेक्स्ट को फ़ाइल-जैसे एक्सप्रेशन और प्रतीकों के साथ मिलाता है, तो इस तरह की खोज करना मुश्किल हो जाता है। उदाहरण के लिए, **पार्क ब्रोडवे** जैसे प्लेटफॉर्म पर, प्रश्न या उत्तरों में अक्सर टेक्स्ट कोड स्निपेट्स या फ़ाइल एक्सप्रेशन के साथ मिश्रित होता है।

लेकिन निश्चित रूप से, इस क्षेत्र में नए उत्पादों की कल्पना की जा सकती है जो प्राकृतिक भाषा खोज और फ़ाइल-आधारित खोज के बीच की खाई को पाट सकें।