

بسمه تعالی

معرفی منابع جایگزین برای پروژه اول بوت کمپ تحلیل داده



آخرین ویرایش

۱۴۰۱/۱۱/۱۱

به دلیل اینکه اگر همه گروه‌ها روی یک سایت اقدام به کروالینگ کنند، ممکن است سایت مورد نظر Down شود، بهتر است دو منبع جایگزین داشته باشیم و بسته به نظر منتور و اعضای گروه، هر گروه یکی از این دو سورس جایگزین را به عنوان منبع اصلی انتخاب کنند.

لازم به تذکر است گروه‌هایی که تا پیش از تغییرات و مشکلات فیدلیو، موفق به دریافت اطلاعات شده بودند، هیچ اجباری در استفاده از این دو سورس ندارند. بلکه به عنوان قسمت اختیاری، می‌توانند اطلاعات این دو منبع را در کنار دیتای فیدلیو تحلیل کنند.

در ضمن به خاطر اینکه برای اتمام پروژه مشکلی متوجه دوستان عزیز نشود، زمان تحویل ۲ روز تمدید و به ۱۷ بهمن‌ماه، ساعت ۲۳:۵۹:۵۹ موکول شد.

## سورس اول

به علت مشکلات پیش آمده برای وبسایت فیدلیو، گروه‌هایی که موفق نشده‌اند تا این لحظه دیتای مناسبی از این سایت دریافت کنند، می‌توانند از وبسایت فودیسم به عنوان منبع جایگزین، استفاده کنند.

همانند سورس فیدلیو، **اولویت با استخراج اطلاعات کافی‌شاپ‌هاست**. در صورتی که اطلاعات کافی‌شاپ‌ها به طور کامل استخراج شدند، می‌توانید به عنوان قسمت اضافه، سراغ رستوران‌ها و سایر موارد بروید.

از آنجایی که اطلاعاتی که فودیسم از هر کافی‌شاپ به دست می‌دهد، شباهت بسیار زیادی به فیدلیو دارد، موارد پیشنهادی برای استخراج، همان مواردی است که در مستند قبلی ذکر شده است.

ایرادی که این سایت دارد، کندی سایت است؛ از این جهت احتمال اینکه وقتی این سایت با تعداد زیادی request روبرو شود، کاملاً down شود وجود دارد!

فلذا نیازمند معرفی سورس دومی هستیم.

## سورس دوم

سورس دومی که به عنوان منبع داده جایگزین استفاده خواهیم کرد، سایت دلینو است. تفاوت دلینو با فیدلیو و فودیسم، این است که مختص رستوران‌هاست و به ندرت می‌توان اطلاعات کافی‌شاپ‌ها را در آن پیدا کنیم.

بنابراین اطلاعاتی که انتظار می‌رود از وبسایت دلینو استخراج کنید، **رستوران‌های هر شهر** است.

به ازای هر رستوران، اطلاعات زیر را می‌توانید در نظر بگیرید:

- شهر
- امتیاز کلی از ۵
- میانگین زمان ارسال
- آدرس
  - آدرس به صورت نوشته (مثلا تهران، سعادت‌آباد)
  - طول و عرض جغرافیایی
- امکان ارسال به خارج از محدوده
- ساعت سرویس‌دهی
- منوی رستوران (غذاهایی که ارائه می‌دهد). به ازای هر غذا، اطلاعات زیر را می‌توانید در نظر بگیرید :
  - نام
  - قیمت
  - مواد اولیه
  - تخفیف