

سلام:))

اولا که کار با فایل‌ها هم برای من هم برای لپ‌تاپ عزیزم بسیار سخت بود. خیلی این پروژه برامون استهلاک داشت راستش.

خب من اول خواستم با استفاده از DB Browser for SQLite فایل‌های یاگو داده شده رو تقریبا extract کنم و بعد به sql تبدیل کنم ولی حجم فایل‌ها بسیار بالا بود و پس از تلاش‌های فراوان نشد. پس به سراغ زبان سلیس پایتون رفتم. با استفاده از توابع کتابخانه‌ای پایتون مثل csv و حتی tsv تونستم به کمی این فایل‌های یاگو رو ببینم. (!) بعد به برنامه نوشتم با پایتون تا بیاد ببینه این فایل‌های یاگو کلا چه رابطه‌هایی رو بررسی کردن. من این کد رو درون فایل زیپ قرار دادم.

بعد که نتیجه‌ی کد ریلیشن‌ها رو دیدم فهمیدم که باید با فایل‌های yagoFacts.tsv و yagoLabels.tsv و yagoTypes.tsv بیشترین کار رو بکنم. کلی گشتم و گشتم و کلی سعی کردم سیاسی فکر کنم. به یه نتیجه‌ای هم رسیدم این وسط: تو ایران حتی غیرسیاسی‌ترین چیزها هم سیاسی محسوب میشن.

بعد به کد پایتون نوشتم که توش بتونم با سرچ کلمه‌هایی مثل "Iran" یا "Iranian" و امثال این‌ها ببینم کلا چه جنس اطلاعاتی توی فایل‌های یاگو هست. خب خوش‌بختانه یاگو از هیچ اطلاعاتی دریغ نکرده بود. خب به برگه برداشتم و شروع کردم به طراحی...

اول از صورت سوال توی کوئرا تقلب کردم و یه جدول politicians طراحی کردم. توی یاگوفکتس گشتم و دیدم مکان مرگ و مکان تولد و جنسیت اکثر سیاستمدارامون اون تو هست. بخاطر همین این ویژگی‌ها هم گذاشتم تو جدول. توی فایل یاگوتایپس هم به سری از اسم رهبران و اسم کسانی که در مجلس بودند و هستند هم وجود داشت. اون‌ها رو هم جزو سیاستمدارها دسته بندی کردم و دوتا ریلیشن به نام‌های افراد مجلس و رهبران هم طراحی کردم که آیدی این دسته از سیاستمدارها رو توی یه ریلیشن نگه دارم. گشتم دیدم یه دسته‌بندی دیگه‌ای هم مثل اسامی آیت‌الله‌ها و کسانی که توی انقلاب شرکت داشتن هم هست. بر کسی پوشیده نیست که آیت‌الله‌ها سیاسی هستند ولی جزو سیاستمدارها دسته‌بندی نمیشن. پس من یه تیبل افراد آیت‌الله طراحی کردم و یه رابطه هم بین سیاستمدارها و آیت‌الله‌ها زدم تا اونایی که توی یاگو توی هر دو دسته قرار می‌گیرن، آیدی‌هاشون تو اون رابطه ثبت بشه. همین کار رو هم برای افرادی که در انقلاب شرکت کرده بودن هم انجام دادم.

از طرف دیگه لیست تمامی احزاب ایران رو درآوردم و با یه ریلیشن به سیاستمدارها وصل کردم. یه مورد دیگه‌ای که پیدا کردم، زندانی‌هایی بود که تو ایران زندانی شده بودن. تو ایران زندانی‌های سیاسی همیشه معروف میشن و با شناختی که داشتم دیدم واقعا هم اکثرشون سیاسی‌ان. یه تیبل سیاسی طراحی کردم. خب هممون می‌دونیم که خیلی از زندانی‌ها به راحتی توی زندان کشته میشن. پس یه ریلیشن هم طراحی کردم و آیدی زندانی‌هایی که تو زندان کشته شدن هم وارد کردم. یه سری اسم سفیرهای مرتبط با ایران هم پیدا کردم. چه کسانی که سفیر ایران در خارج هستن و چه کسانی که سفیر خارجی در ایران هستن یا بودن. یه تیبل طراحی کردم با دو ویژگی اسم سفیرها و برچسب فارسی اسمشون. یه سری اتفاق‌های بودن که توی ایران اتفاق افتادن حالا می‌تونست جنگ باشه، می‌تونست کنفرانس باشه یا رویداد ورزشی یا هرچی. اینو هم از یاگو فکتس پیدا کردم با ریلیشن <happenedIn>. این هم یه تیبل شد برام. یه جدول به نام کشورها ساختم که توش تمام کشورهایی که طی یه اتفاق سیاسی قراره با ایران ریلیشن داشته باشن رو توش قرار بدم. برای کشورها اطلاعات زیادی وجود داشت که خب استفاده کردم و گذاشتم توی دیتابیس. اتفاقی که به‌خاطرش این کشورها رو یه جدول کردم این‌ها بودن:

۱. کشورهایی که ایران با اون‌ها مذاکره کرده بود. ۲. کشورهایی که با ایران مرز سیاسی دارن یعنی همون همسایه‌های ایران. یک جدول هم درست کردم و توش تمام اتفاقی رو که ایران در اون شرکت داشته رو گذاشتم و یه تیبل مرتبط به اون رو هم ساختم و آیدی اتفاق‌ها و کشورهایی که با ایران در اون اتفاق دست داشتن رو قرار دادم.

من از SQLite برای ساختن این دیتابیس استفاده کردم و بعد اون رو به SQL تبدیل کردم. توضیحاتی که فکر می‌کردم روش کارم رو توضیح میده، بالا دادم. نمی‌دونم تونستم کامل توضیح بدم یا نه. تمامی کدهای پایتون در فایل زیپ قرار داده شده. کدها متأسفانه خیلی تمیز نیستن که عذر می‌خوام. پروژه‌ی جالبی بود. (مرسی:)

