

ابتدا عبارت های politician و Iran و ... را با کد پایتون که ضمیمه شده ، در فایل های فکت و تایپ یاگو جستجو کردم و اسامی تمامی سیاست مدارهای ایران را از هر دو فایل استخراج کردم که 520 سیاست مدار شد و از این تعداد آن هایی که برچسب فارسی داشتند را انتخاب کردم که 403 تا شد برای این 403 تا سیاست مدار تمامی خط های فایل های تایپ و فکت یاگو را استخراج کردم که خیلی بیشتر از 2000 خط شد ! سپس همه ی این خط ها را در سه فایل جداگانه EveryLabel-EveryFact-EveryType جدا کردم . تا این جا همه ی کار ها با کد پایتون انجام شد که فرصت نشد برای هر خط آن کامنت بذارم و همانجوری که بود ضمیمه کردم ! سپس پایگاه داده نهایی را طراحی کردم که حدود 13 یا 14 جدول معنی دار اعم از سیاست مدار ، کشور ، دانشگاه و ... شد و حدود 22 تا رابطه بین این جداول بدست آمد که تقریبا همه این جداول با جدول سیاست مدار در ارتباط بودند . کد sql ساخت اسکیمای نیز ضمیمه شده است ، سپس با پایتون به پایگاه داده وصل شدم و فکت ها را از سه فایل که ساختم یکجا به جدول های پایگاه داده Insert کردم که ضمیمه شده است .

در انتها همه اطلاعات در پایگاه داده در دسترس است و به صورت csv از آن dump گرفتم که در پوشه DBFinal هست.