به نام خدا

دانشگاه تهران

پردیس دانشکده‌های فنی

دانشکده برق و کامپیوتر



آزمایشگاه پایگاه‌داده

دستورکار شماره 6

شماره دانشجویی

810196443

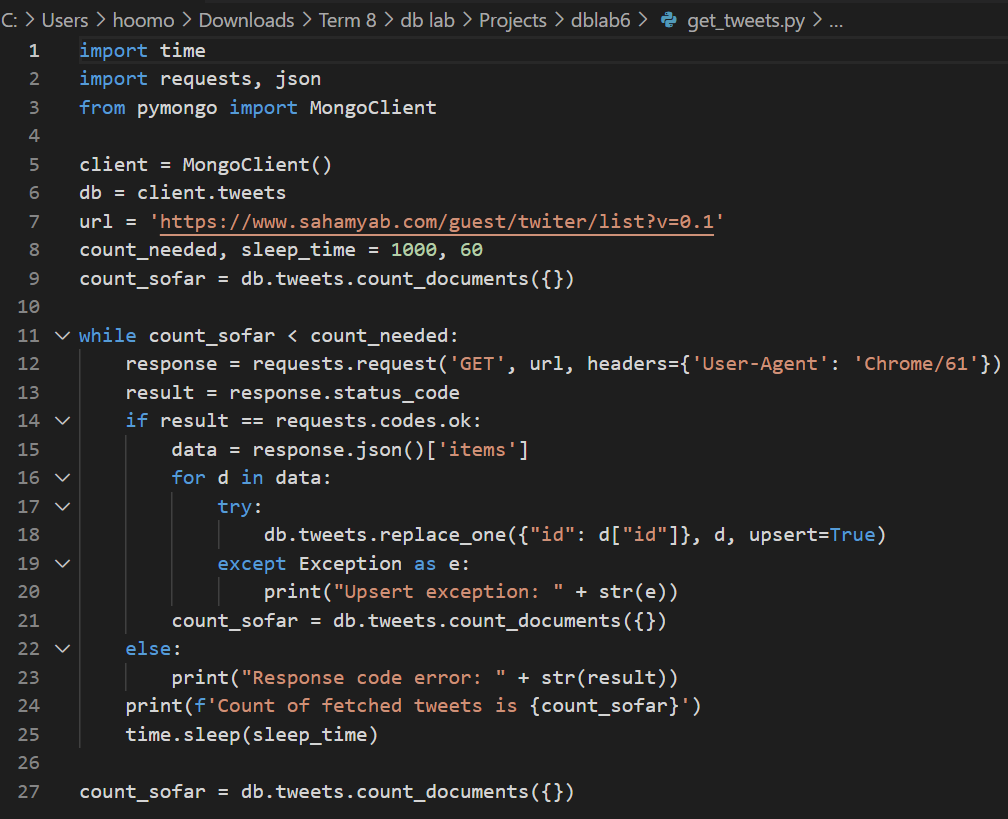
**اردیبهشت 1400**

**هومان چمنی**

# گزارش فعالیت‌های انجام شده

**گام اول:**

داده‌ها به صورت غیردستی و اتومات از طریق یک اسکریپ پایتون به دیتابیس وارد شده‌اند. عکس کد زده شده در زیر قابل مشاهده می‌باشد:



نحوه اطمینان از درج شدن حداقل ۱۰۰۰ ورودی نیز به این صورت است که یک متغیر count\_sofar تعریف شده که نگه‌دارنده تعداد توییت‌هایی است که تا کنون جمع و ذخیره شده. این متغیر در هربار گردش حلقه بروزرسانی شده و شرط خروج از حلقه نیز این است که متغیر ذکرشده حداقل مقدار ۱۰۰۰ را داشته باشد.

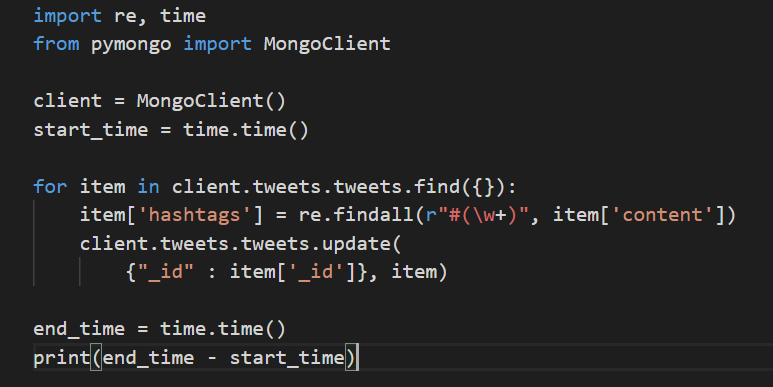
فیلدهایی نیز که اضافه شده‌اند به صورت زیر می‌باشند:

\_id, id, sendTime, sendTimePersian, retwitSendTime, retwitSendTimePersian, retwitSenderName, retwitSenderUsername, retwitSenderProfileImage, senderName, senderUsername, senderProfileImage, content, lastLikeNickName, likeCount, retwitCount, type, scoredPostDate, retwitId, finalPullDatePersian

البته لازم به ذکر است که توییت‌هایی که از جنس retweet نمی‌باشند فیلدهای مربوط به retweet را نیز ندارند.

**گام دوم:**

از کد زیر استفاده کردیم تا هشتگ‌های درون متن را به عنوان یک فیلد جدا اضافه کنیم:



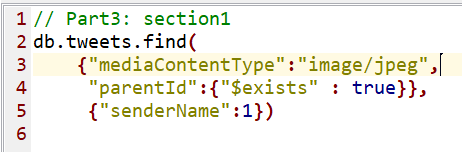
زمان اجرای آن نیز طبق خروجی داده شده مقدار 0.2976 ذکر شده است.

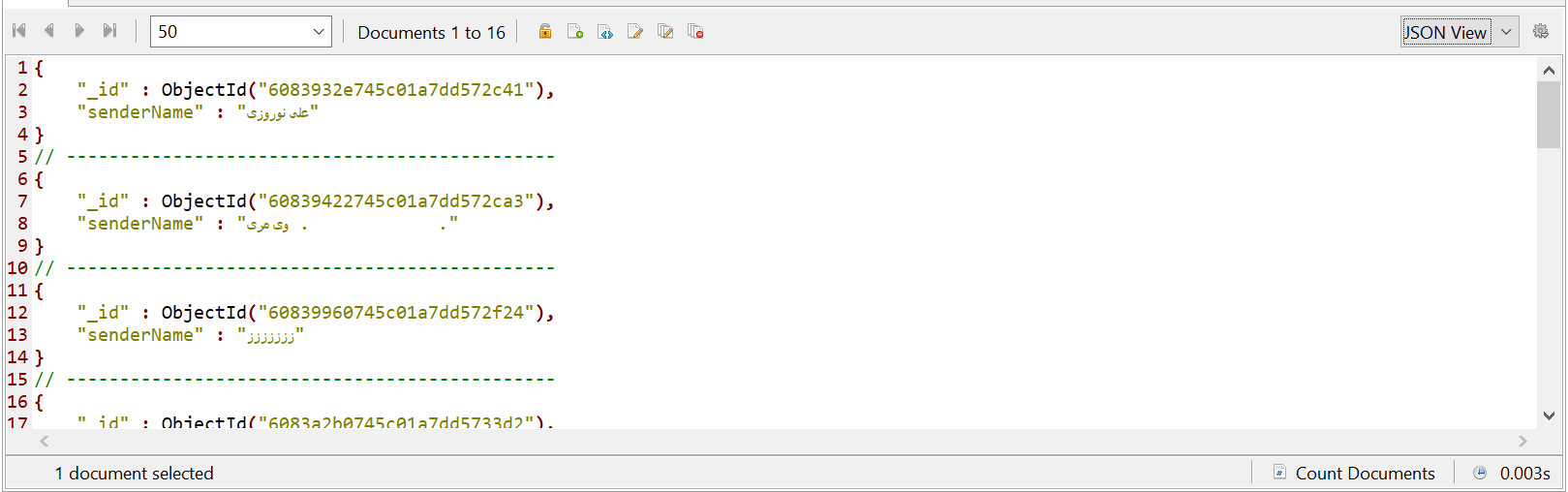
همان‌طور که در تصویر زیر نیز مشاهده‌ می‌شود یک توییت که دارای هشتگ برکت بود به عنوان مثال آپدیت شده است:



**گام سوم:**

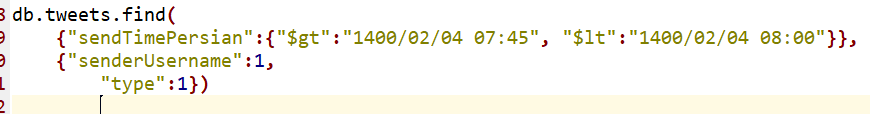
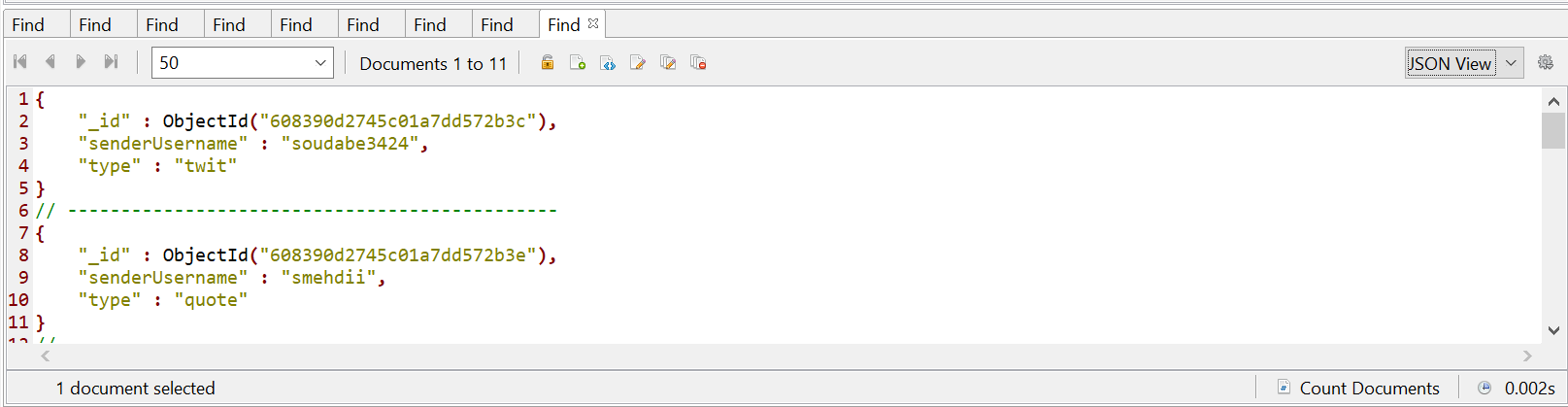
بخش اول:





زمان هم هما‌ن‌طور که در گوشه تصویر بالا معلوم است 0.003 می‌باشد.

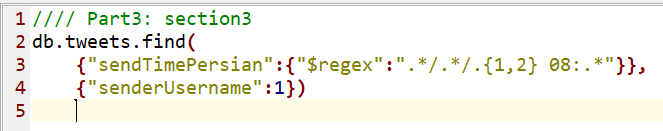
بخش دوم:

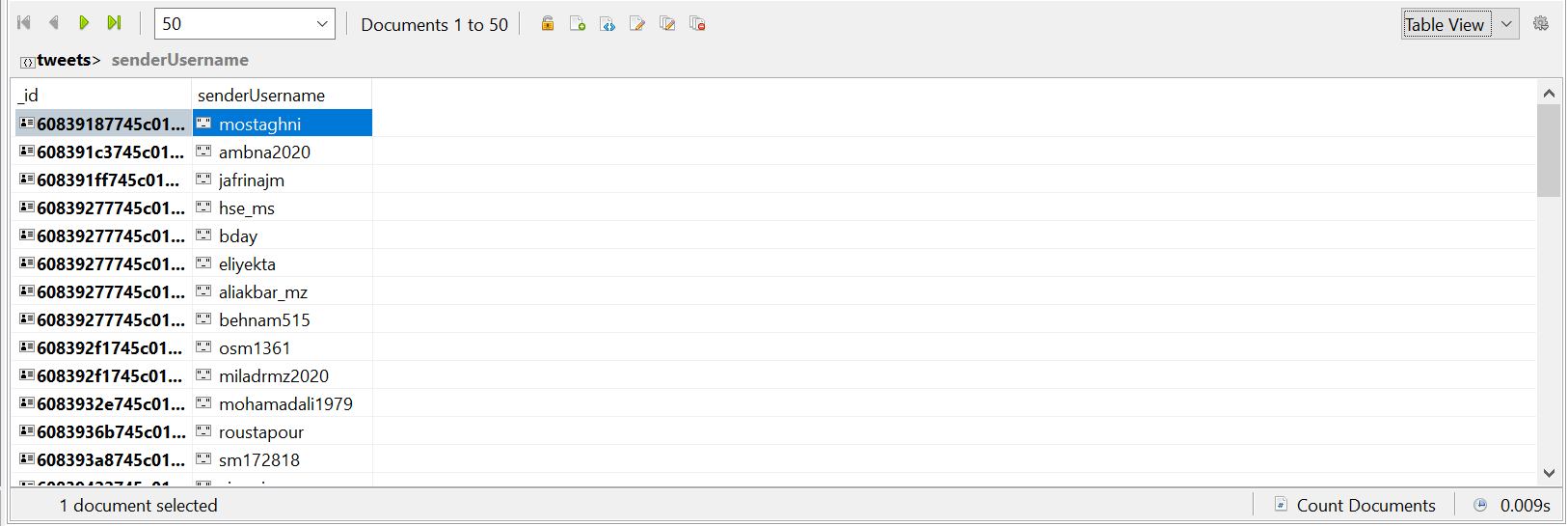
همان‌طور که در شکل بالا مشخص است حدود ۱۱ نفر سحرخیز! داریم که بین بازه ۷.۴۵ تا ۸ صبح توییت کرده‌اند و زمان اجراهم 0.002 می‌باشد.

بخش سوم:

بازه زمانی که شروع آن ۸ صبح است را در نظر گرفته‌ایم.



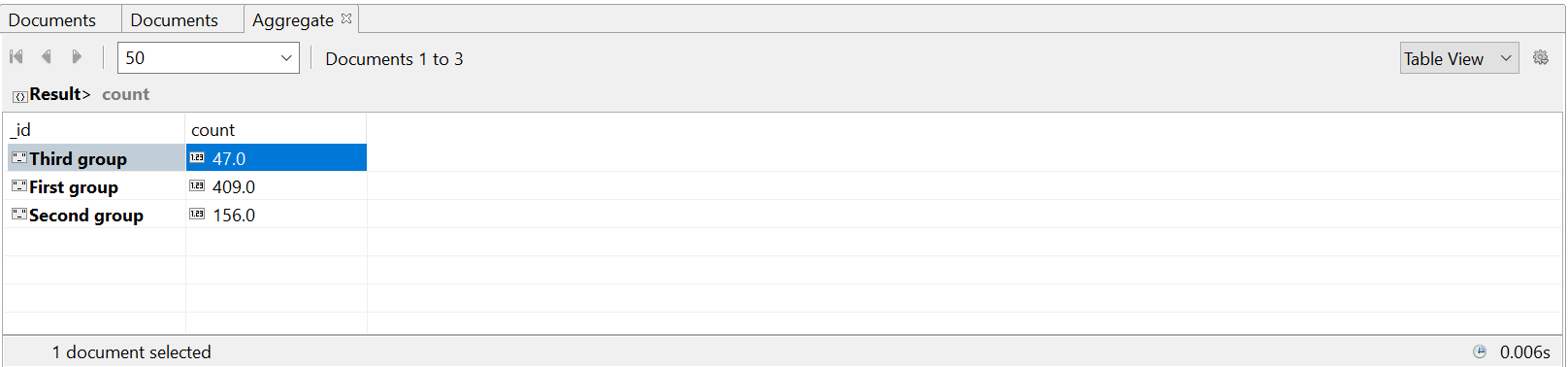
نتیجه ران کردن کووری نیز در زیر قابل مشاهده است. زمان اجرا هم 0.009 ثانیه است:



**گام چهارم:**

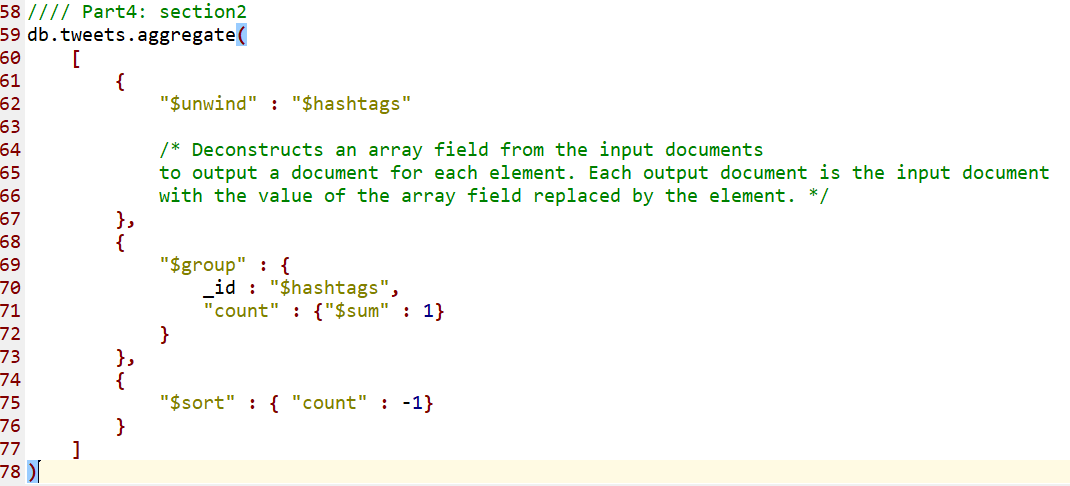
بخش اول:



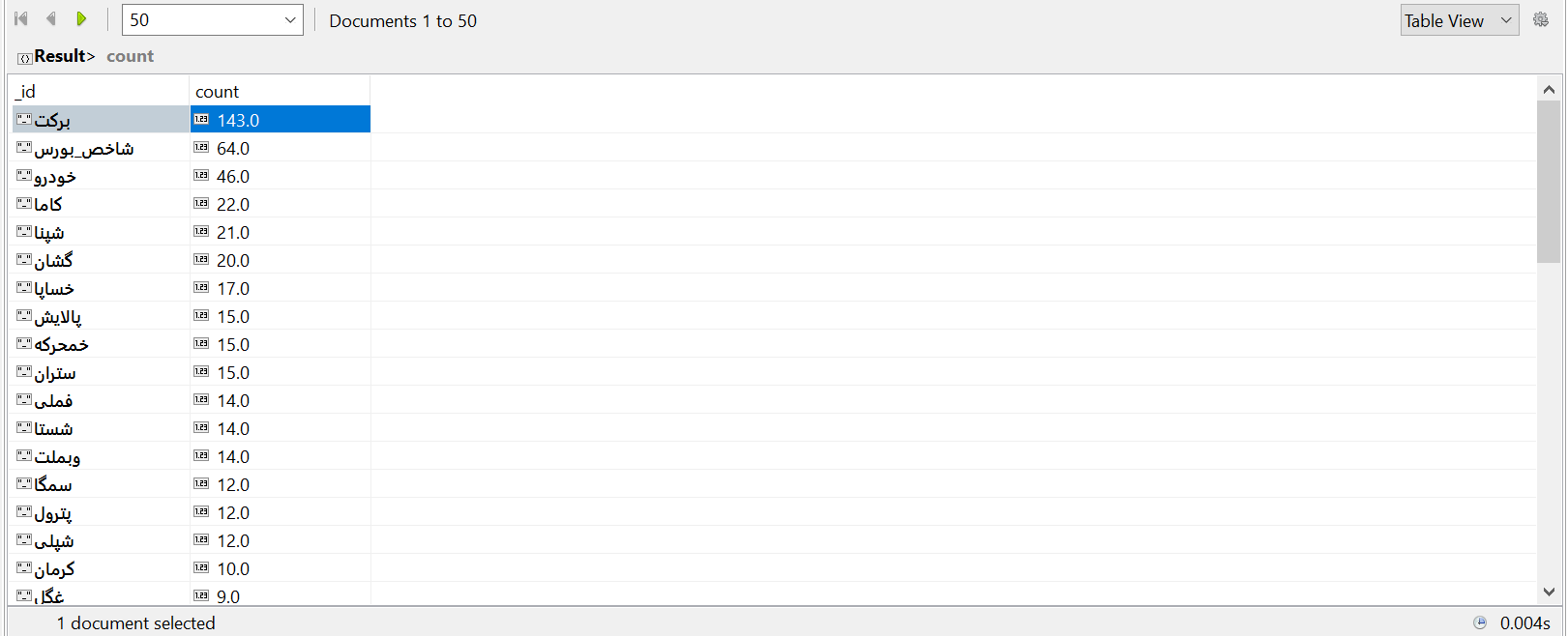


طبق عکس بالا نیز از دسته اول ۴۰۹ کاربر, از دسته دوم ۱۵۶ و از دسته آخر هم فقط ۴۷ کاربر داریم. زمان اجرای کووری نیز 0.006 می‌باشد.

بخش دوم:

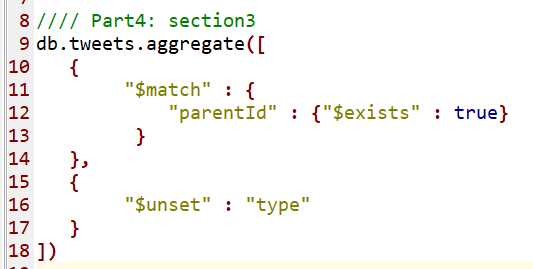


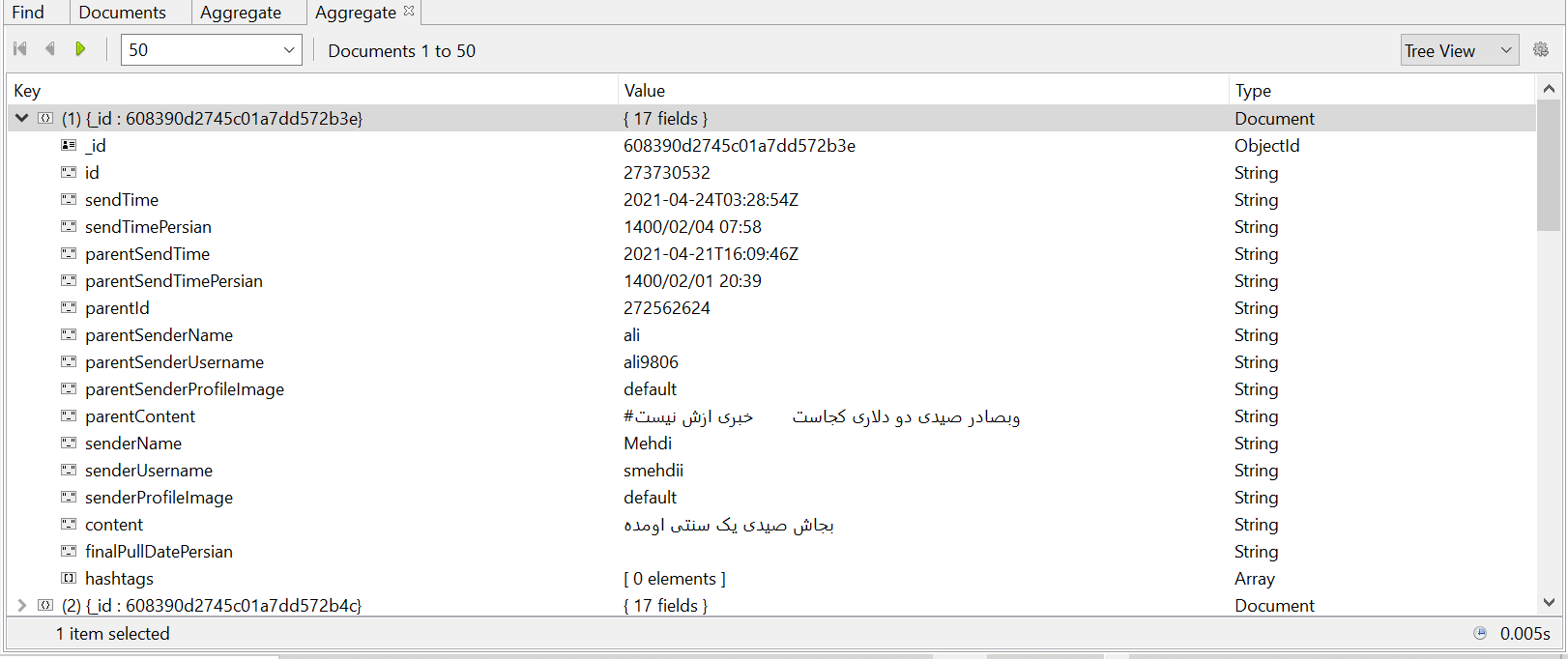
در کووری بالا ابتدا از دستور unwind که توضیح آن زیر آن درج شده استفاده کردیم و در آخر هم سورت طبق تعداد به صورت نزولی انجام داده‌ایم که نتیجه در شکل زیر قابل مشاهده است:



زمان اجرای کووری نیز 0.004 می‌باشد.

بخش سوم:



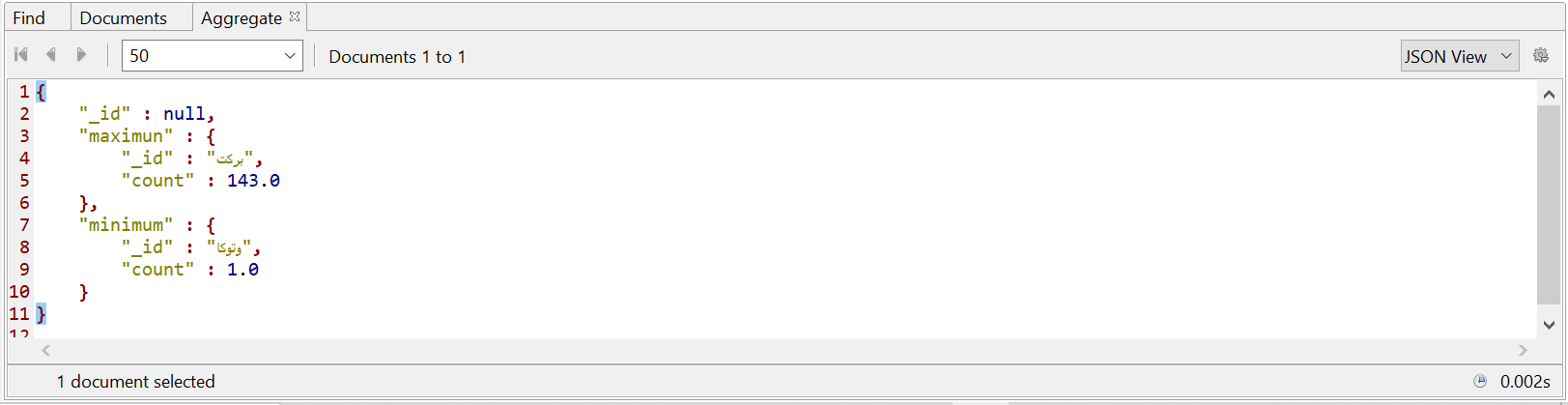


طبق عکس بالا مشاده میشود که دیگر فیلدی به نام type در بین اعضا وجود ندارد. زمان اجرای کووری نیز 0.005 میباشد.

بخش چهارم:

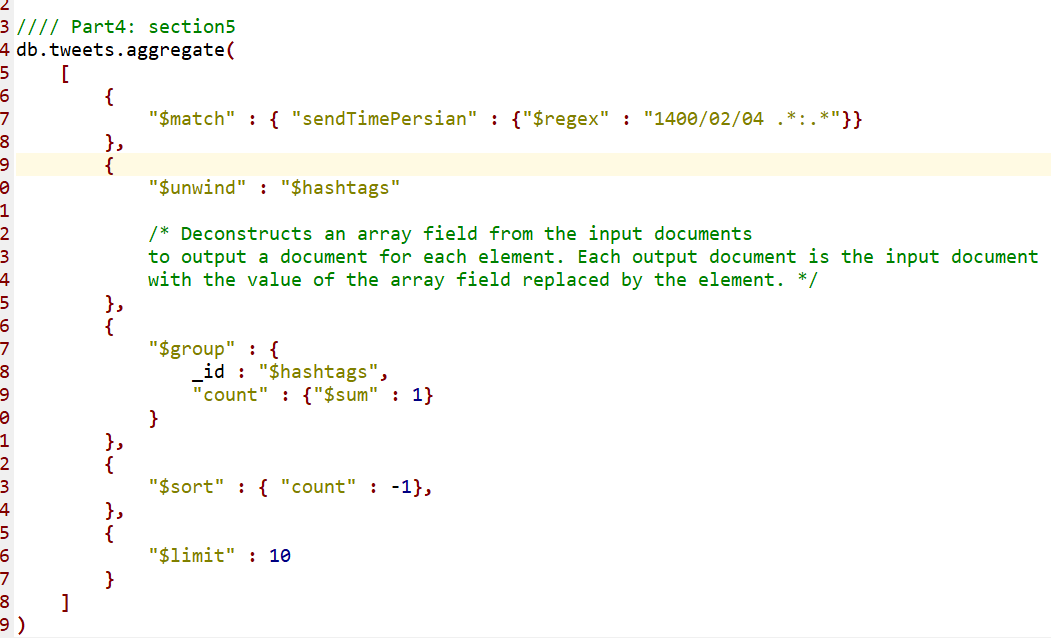
از کووری که در قسمت ۲ زدیم استفاده کردیم با این تفاوت که در آخر آن یک گروه‌بندی دیگر کردیم تا کلیدها با مقادیر ماکس و مین مشخص شوند:



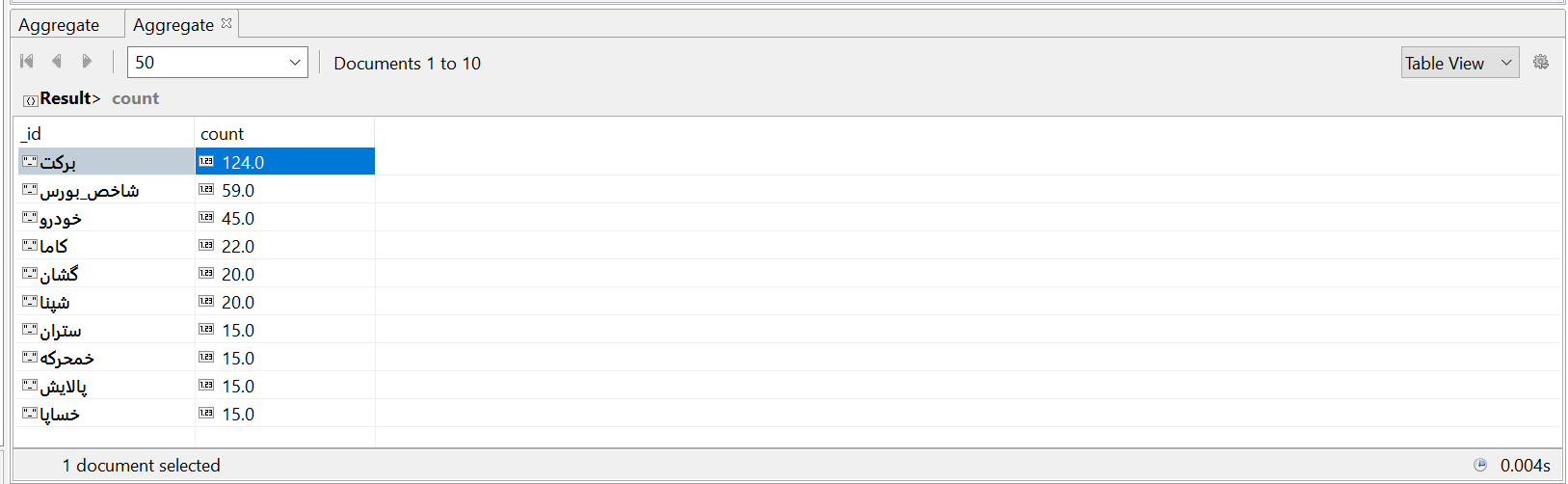


زمان اجرای آن هم طبق عکس بالا 0.002 میباشد.

بخش پنجم:



این کووری مشابه کووری‌های قبلی بوده با این تفاوت که ابتدا صرفا همه عناصری که در تاریخ ۴ اردیبهشت بوده‌اند را در نظر گرفته‌ایم و در آخر هم نتایجی که حاصل شده را صرفا ۱۰ نتیجه اول را خروجی داده‌ایم.

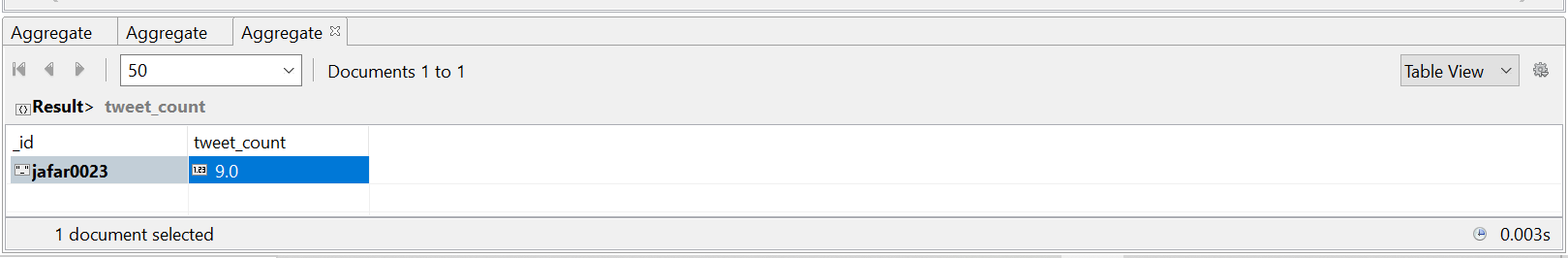


زمان اجرای کووری نیز 0.004 می‌باشد.

بخش ششم:



این کووری مشابه کووری‌ قبلی بوده با این تفاوت که چون فعال‌ترین کاربر را می‌خواهیم محدوده را به ۱ عوض کرده و در گروه‌بندی هم طبق یوزرنیم افراد گروه‌بندی می‌کنیم. نتیجه در شکل زیر قابل مشاهده است:



زمان اجرای کووری نیز 0.003 می‌باشد.

* ز lدرج

# مشکلات و توضیحات تکمیلی

برخی از مواردی که در داکیومنتیشن مونگو وجود داشتند در زدن کووری با 3T با مشکل مواجه شده و نیاز به اندکی تغییر داشتند.

# آنچه آموختم

کار کردن با کووری‌ها در مونگو