2013年12月24日クリスマス

進捗状況

前回ご指摘があった以下の課題をやりました．

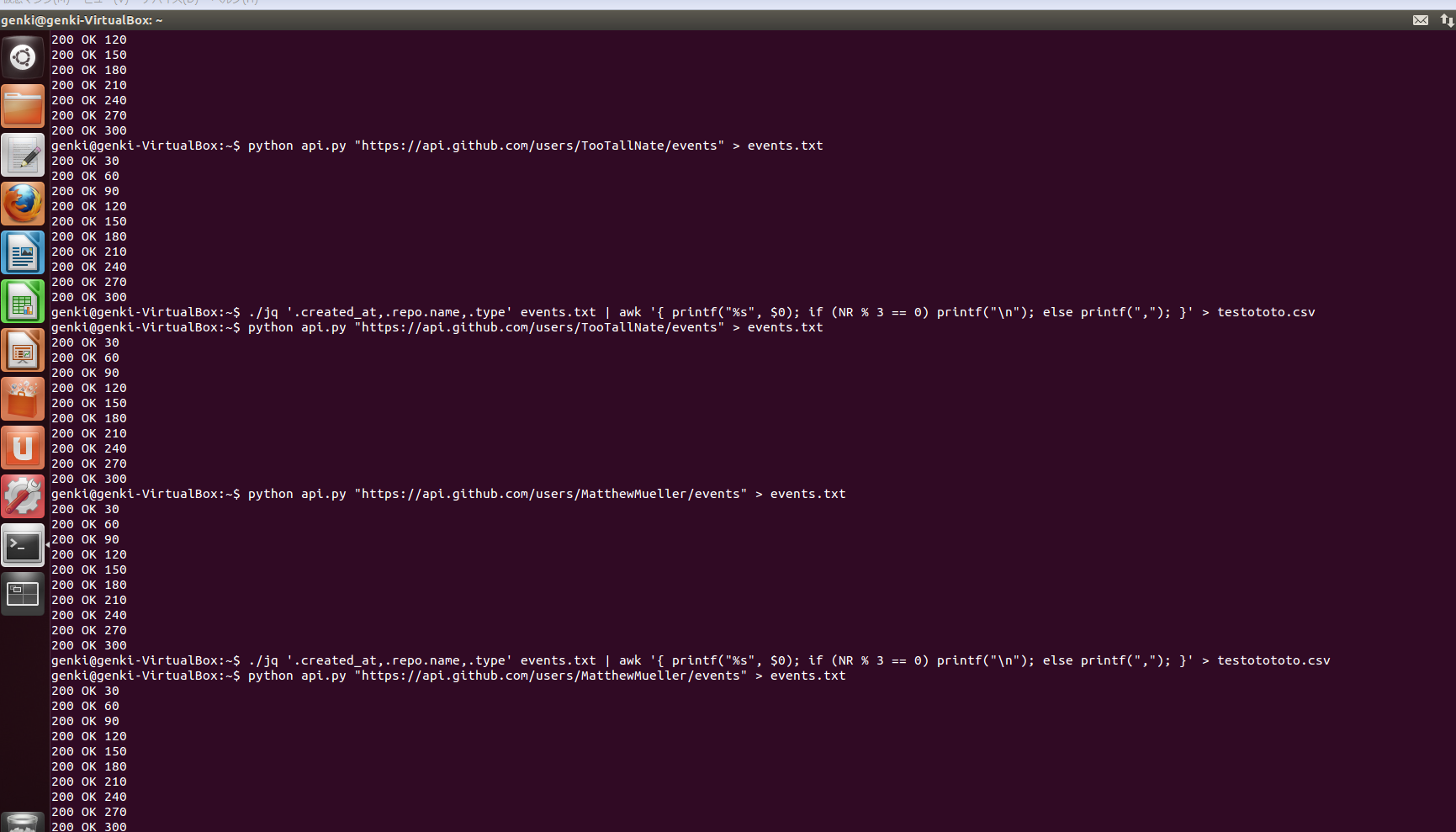
宿題

* どういう形でデータをとれればいいのか，人を指定してデータを取っているか，最終的にどのような形が理想なのかを決めておくといい

まず，先生がGithubに追加してくださった「リポジトリ名も取得する場合」を参照し

リポジトリ名が.repo.nameであることを突き止め、スクリプトを修正しました．

./jq '.created\_at,.repo.name,.type' events.txt | awk '{ printf("%s", $0); if (NR % 3 == 0) printf("\n"); else printf(","); }'を使い，行数を選択しエクセルに表示させました．



　　Guillermo Rauchの活動履歴に加えリポジトリ名も取得している図

後日，複数の人のデータをまとめるため，最初のguille以外の人は、>> events.csvとしてファイルに追記するようにします．

上記の結果

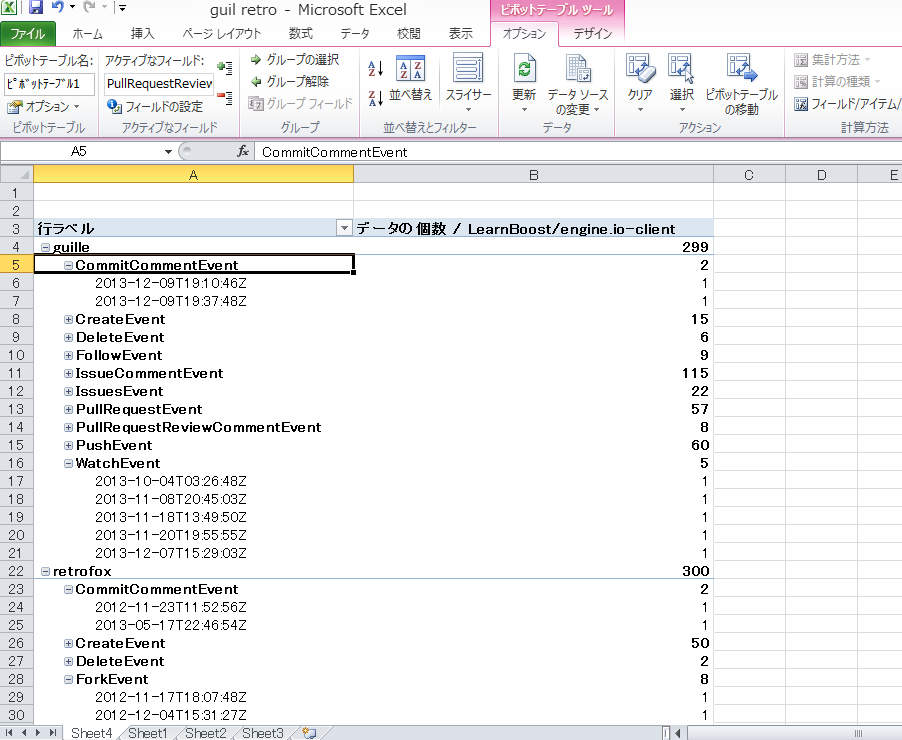
日時，何をしたか，誰が　のみであったデータに

日時，何をしたか，誰が，　どのプロジェクト， を加えることが出来ました．

また，先生が提案してくださった

・あるプロジェクトに関わる二人を指定して，その活動をまとめて取る

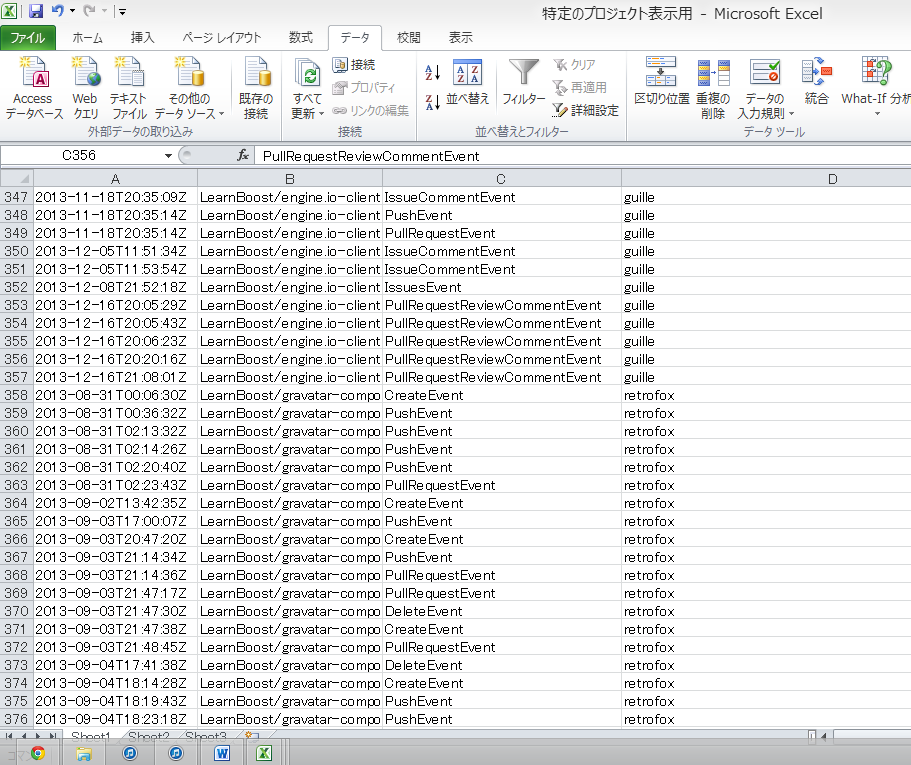
これはLearnBoost で主にメインプログラマのGuillermo Rauch(guille)とサポートとして頻繁に活動しているDamián Suárez　(retrofox)を選び，その活動をまとめとりました．



Guillermo Rauch(guille)とDamián Suárez　(retrofox)のAPIで得たデータを統合している図

またExcelでフィルタをかけ，ある特定のプロジェクト（LearnBoost）だけに絞り表示しました．

先週お話にあった理想とする形に行き着きました．



Guillermo Rauch(guille)とDamián Suárez　(retrofox)のAPIで得たデータを統合したデータからLearnBoostに関するデータのみを表示させた図

宿題

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | CommitComment | Create | Delete | Follow | ... |
| Aさん | 2 | 15 | 6 | 9 |  |
| Bさん |  |  |  |  |  |
| Cさん |  |  |  |  |  |

というような表を作って、

* 主成分分析
* 因子分析
* 自己組織化マップ
* クラスター分析

で分析し、結果を解釈してみる。

そもそも、上の表を自動的に作れるようにしないとつらい。（矢吹が考える）