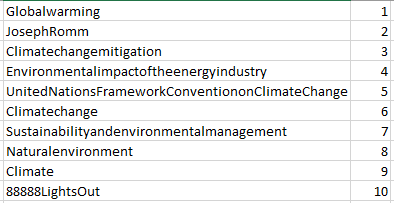
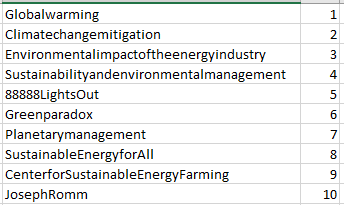
### Query: global warming potential

**TASK 1 TASK2**

If you look to the top results, they are overall similar, only a few here and there but both have the same set of documents.

* Task2 uses normalized tf and idf mainly, that’s the reason I think “Climatechangemitigation” has given higher rank in Task 2, as if see this page,

tf of global in Climatechangemitigation = 125 ; tf of global in JosephRomm = 56

tf of potential in Climatechangemitigation = 25 ; tf of global in JosephRomm = 1

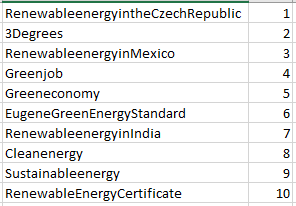
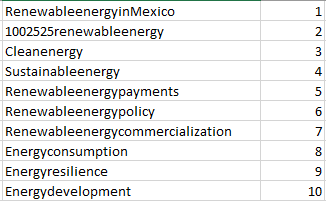
* Lucene normalizes tf by sqrt(frequency), probably that’s why decreasing the factor of term frequency a bit.

The set of top documents is the same in both Tasks, it’s just Task 2 concentrates more on Term Frequency and Lucene (Task1) additionally does the queryNorm which is,

queryNorm = 1 / √sumOfSquaredWeights

whereas, Task 2 considers tf of each query term in the given query only.

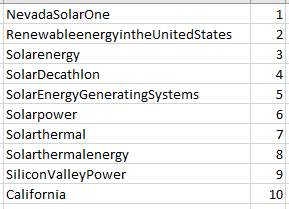
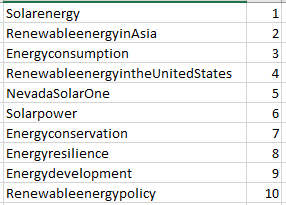
### Query: green power renewable energy

**TASK 1 TASK2**

* It can be seen again that frequency of query terms is more in each of the Task2 results, and its not that Task 2 favors longer documents as “RenewableenergyinMexico” is just a single para document whereas “1002525renewableenergy” is quite long.

### Query: solar energy california

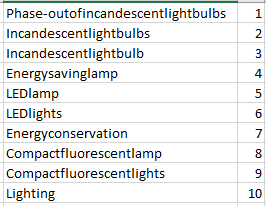
**TASK1 TASK2**

* The set of top results is the same but in different order.
* It has to do with query weights Lucene assigns, as Task2 is dependent on each query term proportional to their tf in query only
* So, Task2 mainly looks for maximum query terms in the same document and then their number of occurrence in the document.

Solarenergy has more frequency of query terms “solar” “energy” than NevadaSolarOne

* It seems like for the given query weightage of term “California” is more than the other two, as its frequency is 4 in NevadaSolarOne and 3 in Solarenergy
* It can be seen when Task1 ranks California at 10, whereas Task2 ranks it at 12.

### Query: light bulb bulbs alternative alternatives

**TASK1 TASK2**

* The top 10 results for both are almost the same, other than only 1 document
* Probably the same root terms are repeated in query, Lucene assigns a similar weightage to each, which is the way Task2 works. That’s why, getting the same results.