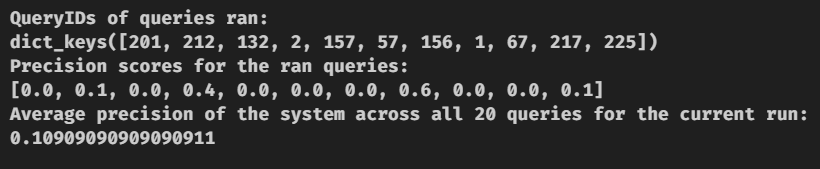
## **How our scoring works**

The *score* function in our notebook takes a query string as well as the matrix of tf-idf values for each term in each document and scores them by counting up the sum of tf-idf values for all terms in the query across their documents. It gives back the top ten documents according to the tf-idf values, ranked descendingly. The *eval* function picks out queries which correspond to at least 15 documents according to the cranqrel file, in order to avoid precision calculations of zero due to none of the documents matching up. It then intersects those IDs with the IDs from the cran.qry file and randomly picks 20 IDs from that intersection. These are the random 20 queries we execute and they are printed when executing the cell containing the *calc\_prec* function. The following is a screenshot from one run of the notebook (bearing in mind that the precision and average precision can vary from run to run due to the stochasticity when randomly picking 20 queries to execute):



For instance, running query number 212 gives a precision of 0.1, meaning 1 of the top-10 scored documents was also found listed in the cranqrel file for queryID 212. Averaging the precision values for all 20 queries run gives a precision of 0.109