Relevance Feedback Report

Programming Assignment 3

Kunal Dhande

Implementation:

- Performance Measures: Precision, Recall and MAP
- First, for the given query, cosine similarity is calculated and top-k elements are returned. Based on these scores, Precision, Recall and MAP scores are calculated.
- Then, current query is submitted to Rocchio algorithm with positive feedback and negative feedback by the user.
- Rocchio algorithm will return the new query based on user feedback
- And this query is again used to retrieve documents with cosine similarity.
- Repeat this for 5 iteration.

For Pseudo Relevance Feedback, instead of user feedback system selects top 3 documents as positive feedback and remaining as negative.

Experimental study - Relevance Feedback

Query 1: kennedy administration pressure on ngo dinh diem to stop suppressing the buddhists .

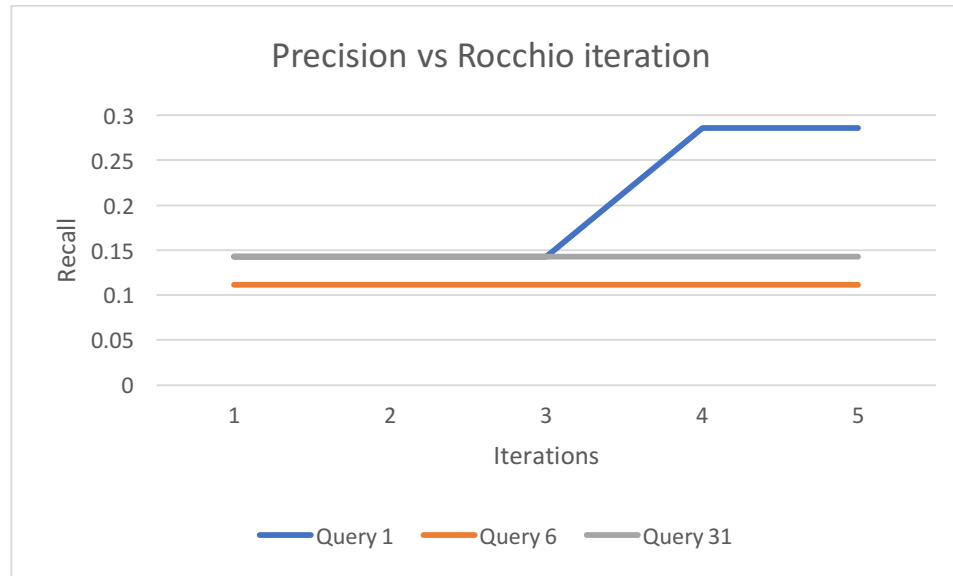| MAP	Recall	Precision	pos_feed_back	neg_feed_back	new_query
0.035714286	0.142857143	0.142857143	[374, 398, 304, 380]	[347, 492, 318, 543]	{'the': 37.0375, 'of': 12.3375, 'a': 13.2375, 'to': 14.425, 'diem': 6.8125, 'and': 11.7, 'that': 6.6, 's': 10.5, 'buddhist': 7.9875, 'in': 11.9625}
0.071428571	0.142857143	0.142857143	[374, 304, 380]	[318, 347, 492, 398]	{'the': 77.38749999999999, 'of': 25.724999999999998, 'a': 27.4375, 'to': 27.9125, 'diem': 13.9625, 'and': 24.4, 'that': 14.075, 's': 23.2125, 'buddhist': 15.912500000000001, 'in': 24.424999999999997}
0.071428571	0.142857143	0.142857143	[304, 268, 326]	[492, 398, 374, 347, 380, 318]	{'the': 101.86249999999998, 'of': 38.425, 'a': 39.3875, 'to': 37.362500000000004, 'diem': 15.2625, 'and': 33.15, 'that': 18.5, 's': 29.0625, 'buddhist': 17.2375, 'in': 32.824999999999996}
0.166666667	0.285714286	0.285714286	[304, 268, 326]	[318, 374, 347, 380, 398]	{'the': 125.58249999999998, 'of': 50.864999999999995, 'a': 51.097500000000004, 'to': 46.612500000000004, 'was': 19.157500000000002, 'and': 41.699999999999996, 'that': 22.88, 's': 34.7225, 'buddhist': 18.3275, 'in': 40.894999999999996}
0.166666667	0.285714286	0.285714286	[304, 268, 326]	[318, 374, 347, 380, 398]	{'the': 149.30249999999998, 'of': 63.30499999999999, 'a': 62.8075, 'to': 55.862500000000004, 'was': 26.527500000000003, 'and': 50.24999999999999, 'christine': 23.25, 'that': 27.259999999999998, 's': 40.38249999999999, 'in': 48.964999999999996}

Query 6: ceremonial suicides committed by some buddhist monks in south viet nam and what they are seeking to gain by such acts .

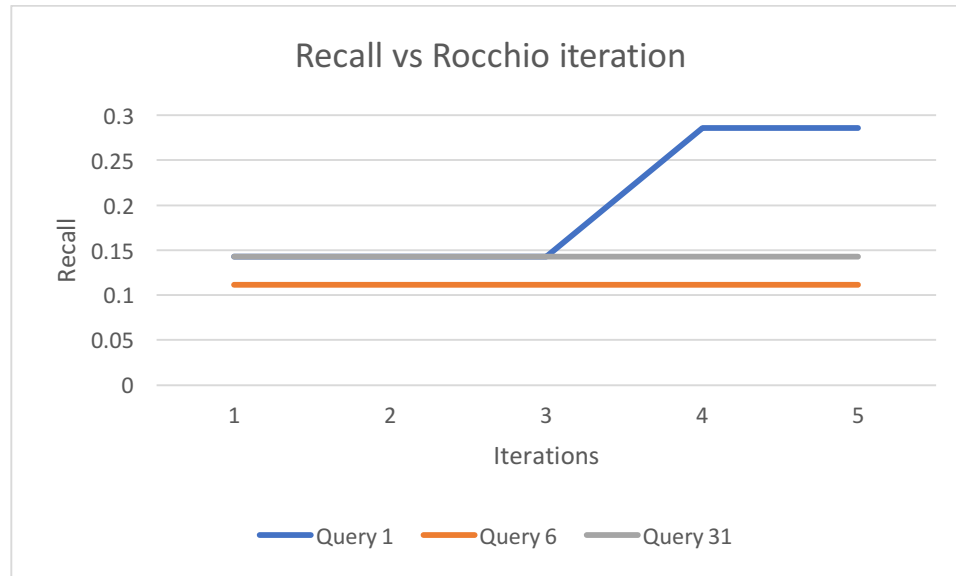| MAP	Recall	Precision	pos_feed_back	neg_feed_back	new_query
0.037037037	0.111111111	0.111111111	[304, 398, 402, 380, 374]	[347, 212, 517, 318]	{'government': 8.025, 'the': 57.0375, 'of': 18.974999999999998, 'a': 18.7125, 'and': 19.675, 'to': 20.837500000000002, 'that': 7.5375000000000005, 's': 15.524999999999999, 'buddhist': 10.5625, 'in': 17.5}
0.027777778	0.111111111	0.111111111	[304, 398, 402]	[517, 212, 374, 347, 380, 318]	{'government': 16.15, 'the': 125.88749999999999, 'of': 42.349999999999994, 'a': 37.0375, 'and': 40.525, 'to': 43.737500000000004, 'that': 15.7375, 's': 30.9, 'buddhist': 20.1125, 'in': 34.825}
0.027777778	0.111111111	0.111111111	[304, 398, 402]	[517, 212, 374, 347, 380, 318]	{'government': 24.275, 'the': 194.73749999999998, 'of': 65.725, 'a': 55.362500000000004, 'and': 61.375, 'to': 66.63750000000002, 'that': 23.9375, 's': 46.275, 'buddhist': 29.6625, 'in': 52.150000000000006}
0.027777778	0.111111111	0.111111111	[304, 398, 402]	[517, 212, 374, 347, 380, 318]	{'government': 32.4, 'the': 263.5875, 'of': 89.1, 'a': 73.68750000000001, 'and': 82.225, 'to': 89.53750000000002, 'that': 32.1375, 's': 61.650000000000006, 'buddhist': 39.2125, 'in': 69.47500000000001}
0.027777778	0.111111111	0.111111111	[304, 398, 402]	[517, 212, 374, 347, 380, 318]	{'government': 40.525, 'the': 332.4375, 'of': 112.475, 'a': 92.01250000000002, 'and': 103.07499999999999, 'to': 112.43750000000003, 'that': 40.337500000000006, 's': 77.025, 'buddhist': 48.762499999999996, 'in': 86.80000000000001}

Query 31: leaders which figure in discussions of the future of the west german chancellorship .

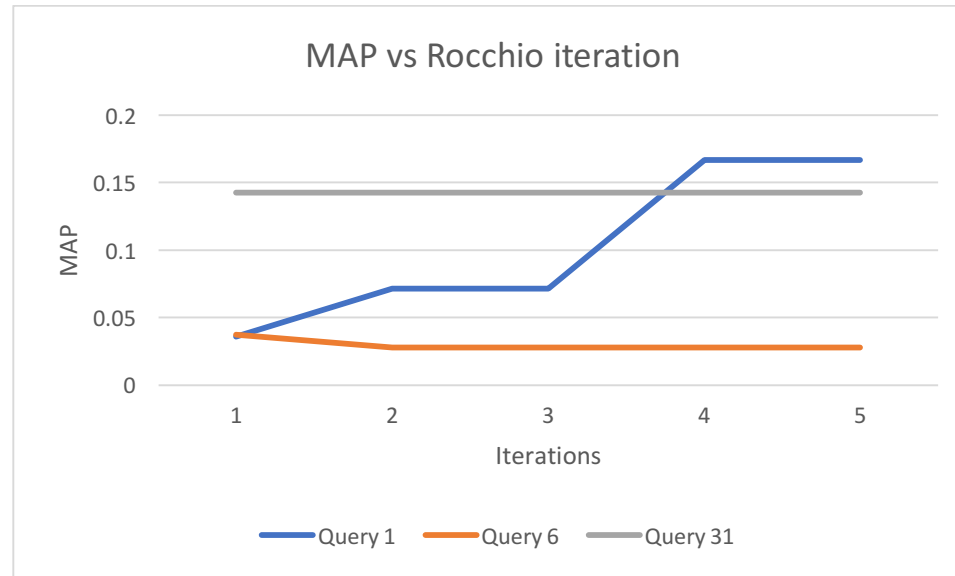| MAP	Recall	Precision	pos_feed_back	neg_feed_back	new_query
0.142857143	0.142857143	0.142857143	[47, 72, 100]	[10, 284, 373, 247]	{'adenauer': 4.0, 'the': 22.2875, 'of': 14.4125, 'a': 7.4875, 'to': 6.9125, 'he': 3.9125, 'for': 4.825, 'and': 6.2625, 's': 8.55, 'in': 6.7}
0.142857143	0.142857143	0.142857143	[47, 72, 100, 10]	[284, 373, 247]	{'the': 43.925000000000004, 'of': 27.400000000000002, 'a': 13.9375, 'to': 12.6875, 'he': 7.3625, 'was': 7.075, 'for': 8.7625, 'and': 12.612499999999999, 's': 16.875, 'in': 12.737499999999999}
0.142857143	0.142857143	0.142857143	[47, 72, 100, 10]	[284, 373, 247]	{'the': 65.56250000000001, 'of': 40.3875, 'a': 20.3875, 'and': 18.9625, 'he': 10.8125, 'was': 10.462499999999999, 'for': 12.7, 'to': 18.4625, 's': 25.2, 'in': 18.775}
0.142857143	0.142857143	0.142857143	[47, 72, 100, 10]	[284, 373, 247]	{'the': 87.20000000000002, 'of': 53.375, 'a': 26.8375, 'and': 25.3125, 'he': 14.2625, 'was': 13.849999999999998, 'for': 16.6375, 'to': 24.237499999999997, 's': 33.525000000000006, 'in': 24.8125}
0.142857143	0.142857143	0.142857143	[47, 72, 100, 10]	[284, 373, 247]	{'the': 108.83750000000002, 'of': 66.3625, 'a': 33.2875, 'and': 31.6625, 'he': 17.7125, 'was': 17.237499999999997, 'for': 20.575, 'to': 30.012499999999996, 's': 41.85000000000001, 'in': 30.85}

Precision vs Rocchio iteration:

Precision vs Rocchio iteration

Recall vs Rocchio iteration:

Recall vs Rocchio iteration

MAP vs Rocchio iteration

From above graphs, we can say that Query 1 improves precision, recall and MAP scores after some iterations of Rocchio algorithm.

Query Drift –

With serval iterations of document retrieval and Rocchio algorithm, there is possibility of query drift. This results into more relevant results.

In above experiment, we can say that for Query 1, there is query drift. And it results into improved precision, recall and MAP scores after some iterations. But, other Queries, viz., 6 and 31, has no change in documents retrieved.

Extra Credit – Pseudo Relevant Feedback

Query 1: kennedy administration pressure on ngo dinh diem to stop suppressing the buddhists .

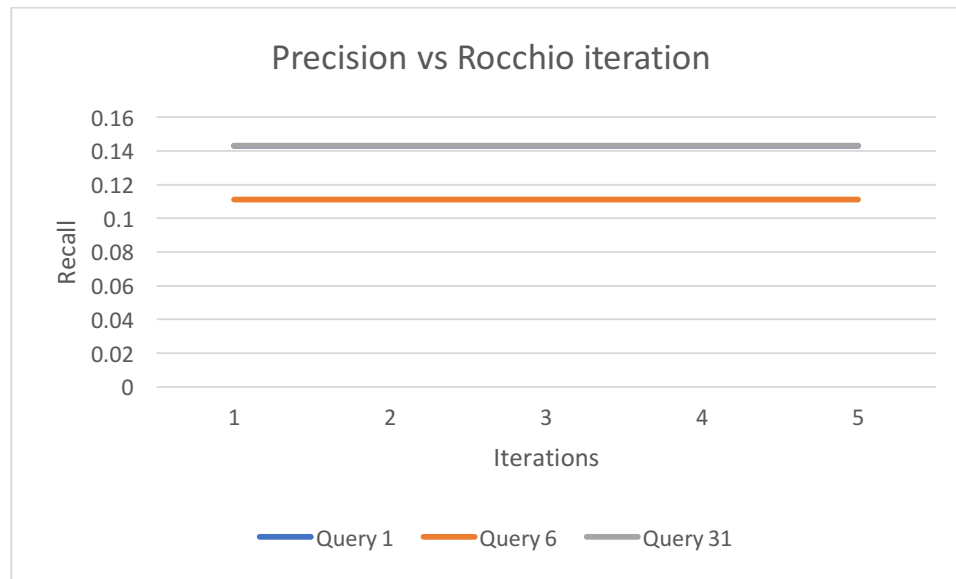| MAP	Recall	Precision	pos_feed_back	neg_feed_back	new_query
0.035714286	0.142857143	0.142857143	[374, 398, 318]	[304, 347, 492, 543]	{'government': 4.4125, 'the': 26.275, 'of': 8.5625, 'a': 9.975, 'to': 14.45, 'diem': 5.3625, 'and': 7.275, 's': 8.7875, 'buddhist': 5.725, 'in': 9.3}
0.142857143	0.142857143	0.142857143	[304, 374, 347]	[318, 398, 492, 543]	{'the': 58.449999999999996, 'of': 19.7125, 'a': 18.512500000000003, 'to': 25.599999999999998, 'diem': 10.3, 'and': 15.7, 'that': 8.587499999999999, 's': 19.012500000000003, 'buddhist': 10.299999999999999, 'in': 18.3125}
0.142857143	0.142857143	0.142857143	[304, 374, 347]	[318, 398, 492, 543]	{'the': 90.62499999999999, 'of': 30.862499999999997, 'a': 27.050000000000004, 'to': 36.74999999999999, 'diem': 15.2375, 'and': 24.125, 'that': 13.599999999999998, 's': 29.237500000000004, 'buddhist': 14.874999999999998, 'in': 27.325}
0.142857143	0.142857143	0.142857143	[304, 374, 347]	[318, 398, 492, 543]	{'the': 122.79999999999998, 'of': 42.012499999999996, 'a': 35.587500000000006, 'to': 47.89999999999999, 'diem': 20.175, 'and': 32.55, 'that': 18.612499999999997, 's': 39.462500000000006, 'buddhist': 19.45, 'in': 36.337500000000006}
0.142857143	0.142857143	0.142857143	[304, 374, 347]	[318, 398, 492, 543]	{'the': 154.975, 'of': 53.162499999999994, 'a': 44.12500000000001, 'to': 59.04999999999999, 'diem': 25.1125, 'and': 40.974999999999994, 'that': 23.624999999999996, 's': 49.68750000000001, 'buddhist': 24.025, 'in': 45.35000000000001}

Query 6: ceremonial suicides committed by some buddhist monks in south viet nam and what they are seeking to gain by such acts.

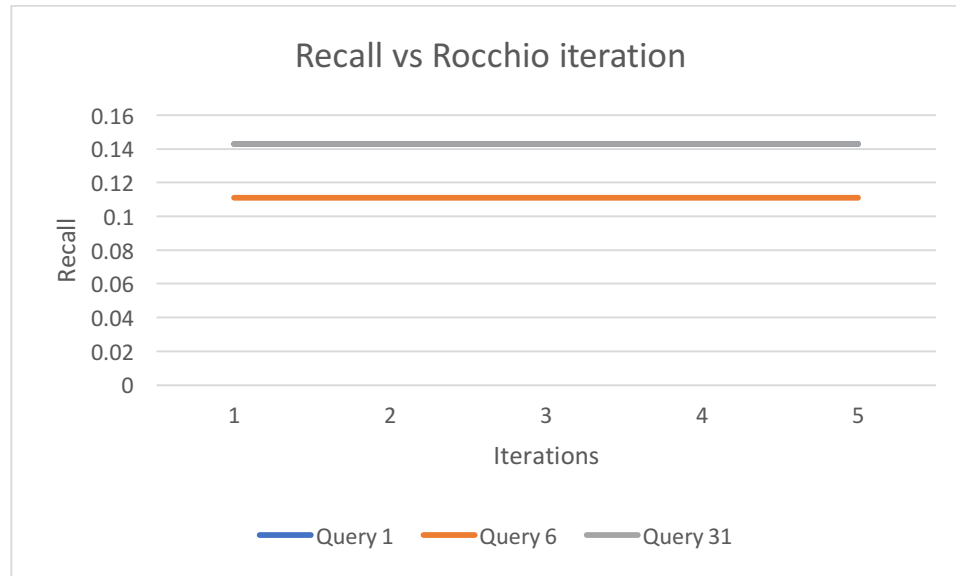| MAP	Recall	Precision	pos_feed_back	neg_feed_back	new_query
0.037037037	0.111111111	0.111111111	[398, 318, 304]	[517, 347, 380, 402, 374, 212]	{'the': 23.200000000000003, 'of': 7.425, 'a': 5.4, 'to': 12.35, 'and': 4.800000000000001, 'by': 4.5, 'that': 4.075, 's': 4.65, 'buddhist': 6.7, 'in': 7.6}
0.027777778	0.111111111	0.111111111	[402, 380, 517]	[304, 374, 347, 318, 398, 212]	{'the': 99.47500000000001, 'of': 34.099999999999994, 'a': 31.7, 'to': 35.25, 'diem': 11.875, 'and': 30.325, 'that': 13.375, 's': 25.799999999999997, 'buddhist': 15.975, 'in': 28.775000000000002}
0.027777778	0.111111111	0.111111111	[402, 380, 517]	[304, 374, 347, 318, 398, 212]	{'the': 175.75000000000003, 'of': 60.77499999999999, 'a': 58.0, 'to': 58.15, 'diem': 20.425, 'and': 55.85, 'that': 22.675, 's': 46.949999999999996, 'buddhist': 25.25, 'in': 49.95}
0.027777778	0.111111111	0.111111111	[402, 380, 517]	[304, 374, 347, 318, 398, 212]	{'the': 252.025, 'of': 87.44999999999999, 'a': 84.3, 'and': 81.375, 'diem': 28.975, 'to': 81.05000000000001, 'that': 31.974999999999998, 's': 68.1, 'buddhist': 34.525, 'in': 71.125}
0.027777778	0.111111111	0.111111111	[402, 380, 517]	[304, 374, 347, 318, 398, 212]	{'the': 328.29999999999995, 'of': 114.12499999999999, 'a': 110.6, 'and': 106.9, 'diem': 37.525, 'to': 103.95000000000002, 'that': 41.27499999999999, 's': 89.25, 'buddhist': 43.8, 'in': 92.3}

Query 31: leaders which figure in discussions of the future of the west german chancellorship .

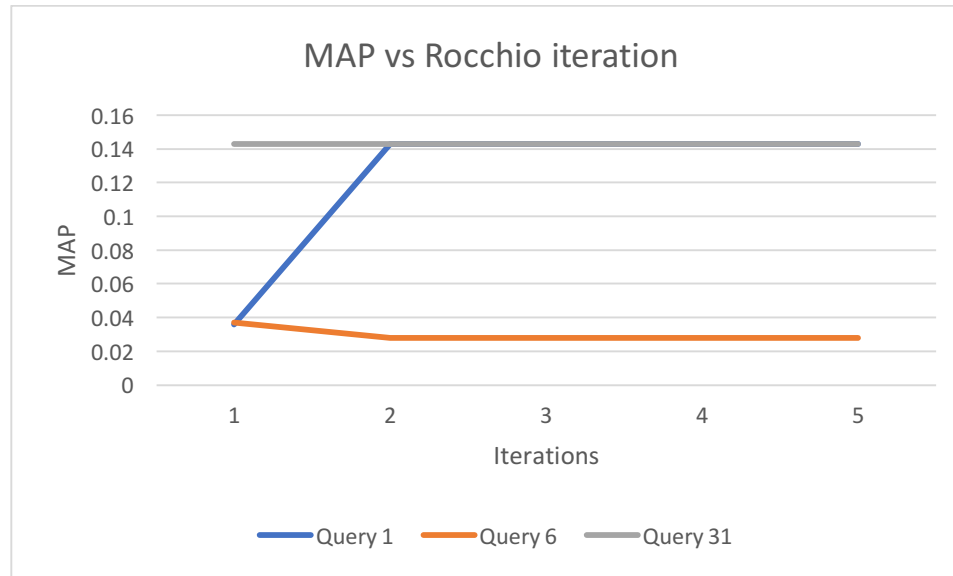| MAP	Recall	Precision	pos_feed_back	neg_feed_back	new_query
0.142857143	0.142857143	0.142857143	[47, 72, 100]	[10, 373, 247, 284]	{'adenauer': 4.0, 'the': 22.2875, 'of': 14.4125, 'a': 7.4875, 'to': 6.9125, 'he': 3.9125, 'for': 4.825, 'and': 6.2625, 's': 8.55, 'in': 6.7}
0.142857143	0.142857143	0.142857143	[47, 72, 373]	[10, 100, 247, 284]	{'the': 45.45, 'of': 25.962500000000002, 'a': 15.55, 'to': 17.275000000000002, 'he': 7.5375, 'for': 9.075, 'and': 13.962499999999999, 'that': 7.3875, 's': 17.962500000000002, 'in': 12.399999999999999}
0.142857143	0.142857143	0.142857143	[47, 72, 373]	[10, 100, 247, 284]	{'the': 68.6125, 'of': 37.5125, 'a': 23.6125, 'to': 27.637500000000003, 'he': 11.1625, 'for': 13.325, 'and': 21.662499999999998, 'that': 11.5125, 's': 27.375000000000004, 'in': 18.099999999999998}
0.142857143	0.142857143	0.142857143	[47, 72, 373]	[10, 100, 247, 284]	{'the': 91.77499999999999, 'of': 49.0625, 'a': 31.674999999999997, 'to': 38.0, 'for': 17.575, 'and': 29.362499999999997, 'that': 15.6375, 's': 36.7875, 'in': 23.799999999999997, 'is': 14.9625}
0.142857143	0.142857143	0.142857143	[47, 72, 373]	[10, 100, 247, 284]	{'the': 114.93749999999999, 'of': 60.6125, 'a': 39.7375, 'to': 48.3625, 'for': 21.825, 'and': 37.0625, 'that': 19.7625, 's': 46.2, 'in': 29.499999999999996, 'is': 18.775}

Precision vs Rocchio iteration:

Recall vs Rocchio iteration:

MAP vs Rocchio iteration

Comparison

Query – 1 :

kennedy administration pressure on ngo dinh diem to stop suppressing the buddhists .

User Feedback			Pseudo Feedback		
Precision	**Recall**	**MAP**	**Precision**	**Recall**	**MAP**
0.14	0.14	0.04	0.14	0.14	0.04
0.14	0.14	0.07	0.14	0.14	0.14
0.14	0.14	0.07	0.14	0.14	0.14
0.29	0.29	0.17	0.14	0.14	0.14
0.29	0.29	0.17	0.14	0.14	0.14

For Query 1, we can say that in case of user feedback, Precision, Recall, and MAP scores are improved after some iterations. But, for pseudo feedback there is no change in these scores.

Query – 6 :

ceremonial suicides committed by some buddhist monks in south viet nam and what they are seeking to gain by such acts .

User Feedback			Pseudo Feedback		
Precision	Recall	MAP	Precision	Recall	MAP
0.11	0.11	0.03	0.11	0.11	0.03
0.11	0.11	0.04	0.11	0.11	0.04
0.11	0.11	0.04	0.11	0.11	0.04
0.11	0.11	0.04	0.11	0.11	0.04
0.11	0.11	0.04	0.11	0.11	0.04

For Query 6, we can say that in case of user feedback and pseudo feedback, both have same scores and there is no change in scores after iterations

Query – 31 :

ceremonial suicides committed by some buddhist monks in south viet nam and what they are seeking to gain by such acts .

User Feedback			Pseudo Feedback		
Precision	**Recall**	**MAP**	**Precision**	**Recall**	**MAP**
0.14	0.14	0.14	0.14	0.14	0.14
0.14	0.14	0.14	0.14	0.14	0.14
0.14	0.14	0.14	0.14	0.14	0.14
0.14	0.14	0.14	0.14	0.14	0.14
0.14	0.14	0.14	0.14	0.14	0.14

For Query 31, we can say that in case of user feedback and pseudo feedback, both have same scores and there is no change in scores after iterations