

Experiments and Results on 24 May, 2018

1. Removing Others [~0]: Role Others were leading to High Accuracy because of their abundance. These are also not relevant for our purpose. Exclude those results as well as LOC_Victim. This role is very less frequent and we will be removing it.

2. Removing Stop Words and Punctuation[~Stop]: Since the stop words were present in the neighbourhood of all the Roles and Entities, all the representations were having them also they are insignificant.

3. Removing Duplicate Results [~DupRes]: There were multiple mentions of the same entity tag pair in result.

4. Ranging K from 1 to 15: We tried to range K from 1 to 15 to see the variation.

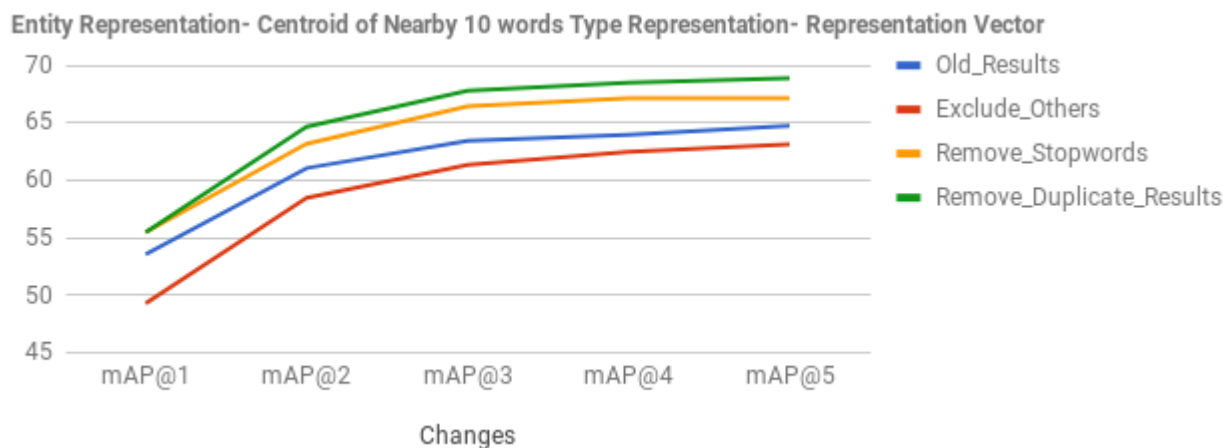
Order of application

Old Setting -> Remove others from Results -> Remove Stop words -> Remove Duplicate Results

Configuration E-V-C-N10 + T-V-LC :

Entity Representation- Centroid of Nearby 10 words
Type Representation- Representation Vector

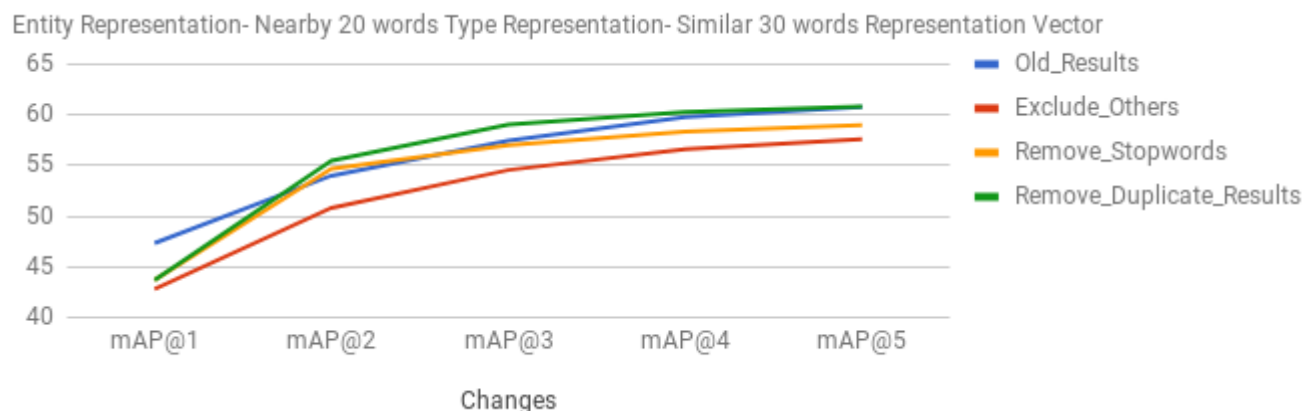
Changes	mAP@1	mAP@2	mAP@3	mAP@4	mAP@5
Old_Results	53.59	61.08	63.43	63.96	64.76
Exclude_Others	49.33	58.48	61.37	62.49	63.14
Remove_Stopwords	55.5	63.2	66.44	67.16	67.14
Remove_Duplicate_Results	55.5	64.68	67.82	68.48	68.91



Configuration E-W-N10 + T-V-SW30 :

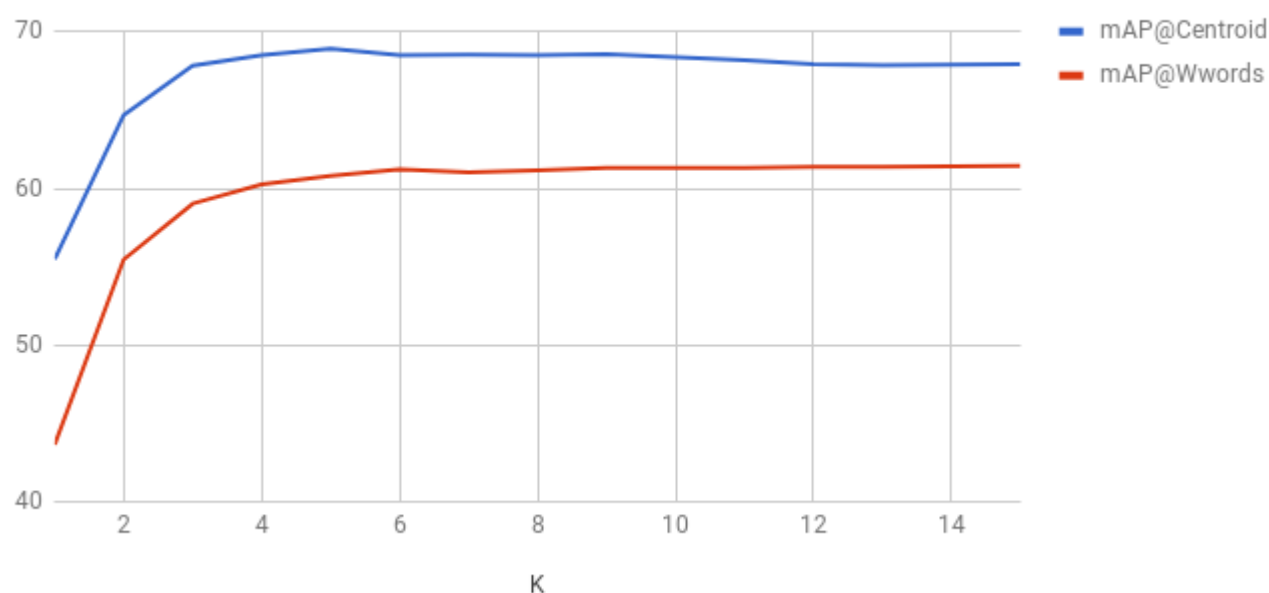
Entity Representation- Nearby left 10, right 10 words
Type Representation- 30 Similar words to Type Representation Vector

Changes	mAP@1	mAP@2	mAP@3	mAP@4	mAP@5
Old_Results	47.33	53.99	57.46	59.76	60.77
Exclude_Others	42.78	50.81	54.55	56.59	57.62
Remove_Stopwords	43.67	54.67	57	58.36	58.97
Remove_Duplicate_Results	43.67	55.47	59.03	60.26	60.81



4. As we increase K from 1 – 15 the performance becomes saturated and starts to go down.

Variation of mAP with K



Summary:

1. Excluding xxx_Others and LOC_Event reduced the performance score by **~3-4%**
2. Removing stop words have given better representative words increasing the overall performance in all configurations. by **~4-5%** in centroid representation and **~2-3%** in word representation.
3. Removing duplicate results have also improved the performance, adding more relevant results. Though the performance is not increased much because **24%** entities have multiple mentions in our corpus.

New Representative words for Role after removing stop words

There were so many stop words in the representative words of role. After Removing Stop words, we were able to get few good words.

LOC_Event:

'298.8', 'friday', 'one', 'bomb', 'near', 'blast', 'monday', 'two', 'waste-papers', 'wednesday', 'riyas', 'people', 'brought', 'area', 'killed', 'blasts', 'attack', 'another', 'first', 'said', 'asimanand', 'time', 'outside', 'explosion', 'day', 'just', 'khallikote', 'since', 'ttp-pakistan', 'went'

LOC_Accused

'asimanand', 'shabari', 'convenership', 'riyas', '9497990026', 'kunangudi', '1-18', '0475-2222595', 'bomb', 'kookuntla', 'immadoddin', 'charge-sheet', 'one', 'kalasangra', 'harekrushna', 'saif-ur-rehman', 'muzafarpur', 'sundershan', 'arrested', 'viqar', 'accused', 'brought', 'khayum', 'abubucker', 'charge-sheets', 'gupta', 'shekawal', 'charge', 'sharp-shooter', 'two'

PER_Accused

'accused', 'convenership', 'arrested', 'case', 'shabari', 'charge-sheet', 'sundershan', 'charge', '1-18', 'charge-sheets', '0475-2222595', 'one', 'brought', 'mr.', 'others', 'harekrushna', 'abinav', 'bomb', 'court', '9497990026', 'time', 'two', 'though', 'said', 'monday', 'police', 'several', 'friday', 'earlier', 'markandey'

PER_Victim

'one', 'injured', 'near', 'said', 'killed', '0.8741815090179443', 'mr.', 'another', 'maudany', 'blast', 'bahadurganj', 'later', 'others', 'sundershan', 'asimanand', 'people', 'riyas', 'went', 'abubucker', 'several', 'muzafarpur', 'friday', 'monday', 'dead', 'attack', 'however', 'convenership', 'time', 'told', 'home', 'minister'

ORG_Victim

'attack', 'bomb', 'ttp-pakistan', 'monday', 'attacks', 'two', 'one', 'sundershan', 'area', 'khyber-pukhtoonkhwa', 'hinduthat', 'riyas', 'said', 'explosives-laden', 'friday', 'harekrushna', 'near', 'security', 'killed', 'blast', 'wednesday', 'parirakshana', 'another', 'asimanand', 'convenership', 'shabari', '298.8', 'kalasangra', 'several', 'ethiopian-themed'

ORG_Accused

'attacks', 'convenership', 'asimanand', 'shabari', 'two', 'riyas', 'one', 'ttp-pakistan', 'sundershan', 'claimed', 'accused', 'involved', 'earlier', 'attack', 'abubucker', 'said', 'bomb', 'government', 'jehadi/terrorist', 'time', 'blast', 'involvement', 'even', '298.8', 'several', 'ethiopian-themed', '806-page', 'khyber-pukhtoonkhwa', 'group', 'though'