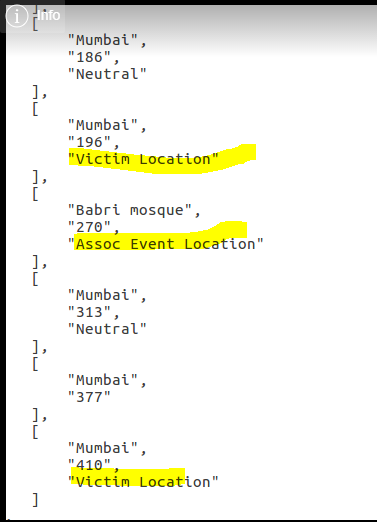
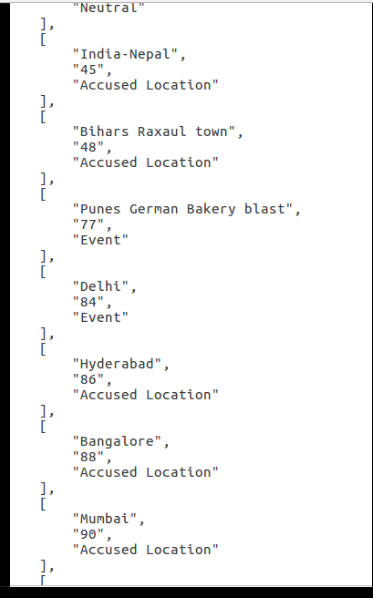
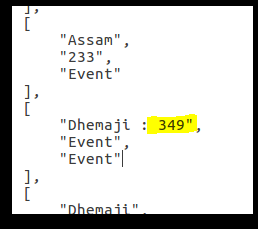
1. Data set consists of 700 documents of bomb blast. 101 events dataset is there but not tagged with Person\_Accused and so on

2. Data set has so many issues, Like missing classes, extra classes, wrong class assignment tagging etc. Format of tagging is also different .. causes issue in generating uniform. Need to go through each doc again. I fear due to bad dataset results will be bad.







**Preprocessing Questions:**

1. Should I test the effect of stop word removal, punctuation mark removal.

2. Lemmatization is needed or not. For SRL may be needed. Should I test the effect of these preprocessing also.

3. There are two ways of TAG generation

New Delhi was attacked.

LOC LOC O O // Taking New and Delhi seperate

LOC O O // Taking New and Delhi same

**Evaluation Question:**

If Others match with Others .. Should we consider that also in evaluation or only classes only. I assume OTHERS should not be considered.