Results

Best scores so far:

Dataset	Precision	Recall	F1
Consolidated MUC	35.131	26.000	29.8851
Unconsolidated MUC	33.33	34.48	33.8983
MUC with GPT labels			

Consolidated:

```
ROLE                  CORRECT      MISSING      OVERPRED    F1
attacker              7            17           4           40.00
target                5            12           13          28.57
instrument            1            7            2           18.18
place                 0            1            5           0.00
TOTAL:                13           37           24          29.89
 -2 & 5 & 20.0 & 14.3 & 16.7 \\ [p r f1 7 gold/5 pred. ]
 -1 & 7 & 28.6 & 25.0 & 26.7 \\ [p r f1 8 gold/7 pred. ]
 0 & 16 & 37.5 & 25.0 & 30.0 \\ [p r f1 24 gold/16 pred. ]
 1 & 8 & 50.0 & 50.0 & 50.0 \\ [p r f1 8 gold/8 pred. ]
 2 & 1 & 0.0 & 0.0 & 0.0 \\ [p r f1 3 gold/1 pred. ]
Precision: 35.1351 Recall: 26.0000 F1: 29.8851
```

Unconsolidated

```
,attacker,target,instrument
attacker,100,0,0
target,0,100,0
instrument,0,0,100
ROLE                  CORRECT      MISSING      OVERPRED    F1
attacker              5            8            9           37.04
target                4            8            4           40.00
instrument            1            3            2           28.57
place                 0            0            5           0.00
TOTAL:                10           19           20          33.90
 -2 & 7 & 28.6 & 33.3 & 30.8 \\ [p r f1 6 gold/7 pred. ]
 -1 & 4 & 0.0 & 0.0 & 0.0 \\ [p r f1 3 gold/4 pred. ]
 0 & 16 & 50.0 & 44.4 & 47.1 \\ [p r f1 18 gold/16 pred. ]
 1 & 0 & 0.0 & 0.0 & 0.0 \\ [p r f1 1 gold/0 pred. ]
 2 & 3 & 0.0 & 0.0 & 0.0 \\ [p r f1 1 gold/3 pred. ]
Precision: 33.3333 Recall: 34.4828 F1: 33.8983
```

```
,attacker,target,instrument,place
attacker,88,11,0,0
target,0,100,0,0
instrument,0,0,100,0
place,0,0,0,100
ROLE              CORRECT      MISSING     OVERPRED     F1
attacker          8            22          19           28.07
target            6            19          22           22.64
instrument        3            9           13           21.43
place             2            1           0            80.00
TOTAL:            19           51          54           26.57
-2 & 9 & 22.2 & 20.0 & 21.1 \\ [p r f1 10 gold/9 pred. ]
-1 & 9 & 11.1 & 20.0 & 14.3 \\ [p r f1 5 gold/9 pred. ]
 0 & 29 & 37.9 & 35.5 & 36.7 \\ [p r f1 31 gold/29 pred. ]
 1 & 4 & 0.0 & 0.0 & 0.0 \\ [p r f1 2 gold/4 pred. ]
 2 & 3 & 0.0 & 0.0 & 0.0 \\ [p r f1 1 gold/3 pred. ]
Precision: 26.0274 Recall: 27.1429 F1: 26.5734
```

44282

SAN SALVADOR , 6 DEC 89 (DOMESTIC SERVICE) -- [NEWS CONFERENCE WITH COLONEL RENE EMILIO PONCE , CHAIRMAN OF THE JOINT CHIEFS OF STAFF OF THE SALVADORAN ARMED FORCES , BY UNIDENTIFIED SALVADORAN AND INTERNATIONAL REPORTERS] [TEXT] [CONTINUED] [REPORTER] COLONEL , WHAT SPECIFIC MEASURES HAVE BEEN TAKEN AGAINST THE MISSILES ? [PONCE] I BELIEVE THAT SPECIFIC MEASURES WE HAVE ADOPTED HAVE GIVEN US GOOD RESULTS . THEY HAVE LAUNCHED FOUR SA-7 'S IN THE ZACATECOLUCA AREA ; HOWEVER , OUR PILOTS HAVE SKILLFULLY AVOIDED THE ROCKETS instrument OR MISSILES , WHICH DID NOT HIT THEIR AIRCRAFT . WE ARE ADOPTING MEASURES THAT ARE GIVING RESULTS . WE MUST ALSO CLARIFY THAT WE FEEL THAT SINCE THE LIGHT AIRCRAFT TRANSPORTING THESE MISSILES DID NOT HAVE A NORMAL LANDING AND THE MISSILES WERE NOT IN THEIR ORIGINAL PACKAGE , IT IS POSSIBLE THAT THEY SUFFERED SOME FUNCTIONAL DAMAGE WHEN THE PLANE CRASHED SOUTH OF ZACATECOLUCA . THANK GOD , [WORDS INDISTINCT] [REPORTER] DO YOU THINK THAT THE SUMMIT OF CENTRAL AMERICAN PRESIDENTS CAN OPEN NEW POSSIBILITIES FOR RESUMING TALKS WITH THE FMLN ? DO YOU THINK THAT THIS SUMMIT COULD BRING ABOUT NEW STEPS TOWARD A DIALOGUE ? [PONCE] YES , I THINK THAT THE SALVADORAN GOVERNMENT SINCE THE BEGINNING OF THE YEAR HAS OPENED THIS POSSIBILITY AND IS WILLING TO TALK ON A PERMANENT BASIS . I THINK THAT AT THIS TIME IT WOULD BE A POSITIVE STEP IN THE POLITICAL FIELD . HOWEVER , IN THE MILITARY FIELD , WHERE THE FMLN HAS ATTACKED US IN AN IRRATIONAL AND VIOLENT WAY IN THE MAIN CITIES , THE ARMED FORCES IS REPLYING , AND WE WILL CONTINUE TO REPLY . WE WILL COMPLETELY DEFEAT THE FMLN . BASED ON THE INFORMATION WE HAVE RECEIVED FROM OUR UNITS , WE HAVE INFLICTED 1,902 DEAD , 1,209 WOUNDED , AND 44 TERRORISTS HAVE DESERTED AND TURNED THEMSELVES IN TO THE AUTHORITIES . WE HAVE CONFISCATED A LARGE NUMBER OF WEAPONS . I WILL GIVE YOU THE QUANTITIES : 25 ANTIAIRCRAFT MISSILES -- ONE RED EYE AND 24 SA-7 'S -- 915 RIFLES , 12 CARBINES , 11 SHOTGUNS , 156 GRENADE AND ROCKET LAUNCHERS , 49 MORTARS , 45 MACHINEGUNS , 41 SUBMACHINEGUNS , 67 REVOLVERS AND PISTOLS , 23 RADIOS , AND A LARGE AMOUNT OF EXPLOSIVES . I MUST STRESS THAT THE FMLN HAS GATHERED ALL THESE WEAPONS WITH SUPPPORT FROM ABROAD . IT IS A LARGE ARSENAL . I MUST ALSO NOTE THAT MANY OF THE WEAPONS WE FOUND HAD BEEN ABANDONED , WHICH MEANS THAT THE FMLN HAS HAD MANY DESERTIONS . WE HAVE ALSO FOUND WEAPONS IN HOUSES , THANKS TO THE COLLABORATION OF THE CIVILIAN POPULATION . WE HAVE ALSO LEARNED THAT THE FMLN HAS KIDNAPPED conflict.attack.n/a A LARGE NUMBER OF YOUTHS HERE IN THE SAN SALVADOR AREA AND MAINLY IN USULUTAN , AS WELL AS IN CHALATENANGO . THIS MEANS THAT THEY ARE TRYING TO FILL IN THE NUMBER OF CASUALTIES THEY HAVE SUFFERED WITH YOUTHS . THIS ALSO MEANS THAT THEIR EFFICIENCY WILL BE VERY LOW . [REPORTER] I DO NOT KNOW IF YOU HAVE SEEN OVER THE PAST FEW DAYS A LEAFLET ISSUED BY THE AIR FORCE , WHICH WAS DROPPED FROM AIRPLANES AND ALSO DISTRIBUTED ON THE STREETS . THE LEAFLET URGES SALVADORAN PATRIOTS TO DEFEND THEIR HOMELAND , THEIR PROPERTIES , AND LIVES . IT ALSO SAYS THAT IF NECESSARY , TERRORISTS MUST BE KILLED , AS WELL AS THEIR INTERNATIONALIST ALLIES . WHAT IS YOUR OPINION OF THIS LEAFLET ? IS IT NOT AN INVITATION TO KILL FOREIGNERS IN THIS COUNTRY ? [PONCE] I DO NOT THINK SO , BECAUSE THE LEAFLETS MENTION INTERNATIONALISTS , WHO ARE FIGHTING ALONG WITH THE FMLN . [CONTINUED]

Other datasets

Wikievents (10 epochs)

Task	Precision	Recall	F1
Role Identification	71.56	67.91	69.15
Role	65.06	62.75	63.38
Coref Role Resolution	71.53	68.98	70.24
Coref Role	66.17	63.81	64.97

RAMS

Window Size 15

Dataset	Precision	Recall	F1
Consolidated MUC	26.0274	27.1429	26.5734

Smaller window size seems to work better for RAMS

Approaches Tried So Far

1. We convert the MUC data to RAMS format to make it compatible and feed it into the UIUC model. We only take one trigger per document(middle trigger in this approach) and use a window of 2 sentences before and 2 sentences after it to get the most relevant context.
2. We do not truncate the document and use all the sentences while continuing to use the middle trigger
3. We take the first trigger and use a window of 2 sentences before and after it because first trigger seems to be a better indicator in TANL.
4. Window Size of 15 instead of 5.
5. Train it to identify one trigger per event and then when you get entities, you pass in the document however and each time specify diff trigger

Multiple triggers for RAMS.
OPtion 1: Split sentences and pass in each
OPtion 2: Try al triggers separately

Input to Algorithm(MUC format):

Consolidation:

https://drive.google.com/file/d/1rxxF_qhoaP_8ChmtUx1VZ8i_VWL0cCeV/view

Unconsolidation:

Here is an example of a single MUC json document from the non-consolidation data.

[20] print(data[0].keys())

dict_keys(['id', 'annotations', 'file_upload', 'drafts', 'predictions', 'data', 'meta', 'created_at', 'updated_at', 'inner_id', 'total_annotations', 'cancelled_annotations', 'total_predictions', 'comment_count', 'unresolved_comment_count', '1

print(data[0])

{'id': 41502, 'annotations': [{'id': 1654, 'completed_by': 1, 'result': [{'value': {'start': 345, 'end': 353, 'text': 'KILLINGS', 'labels': ['trigger']}, 'id': 'ptDxZ3dFgN', 'from_name': 'tag', 'to_name': 'text', 'type': 'labels', 'origin': '

+ Code + Text

Run Through All MUC Events

Go to the colab notebook and show.

RAMS keys are:
sentences, ent_spans, "evt_triggers": [
         [start_index, end_index, [[event type string, 1.0]]]
], "gold_evt_links": [
         [[start_ind_of_trigger, end_ind_of_trigger],
            [start_ind_from_ent_spans, end_ind_from_ent_spans],
            role string (same as thing from ent_spans)
         ]
], source_url, split, doc_key

Here is an example of a single RAMS json document.

Output to Algorithm(RAMS format) and input to UIUC

Consolidation:

"{\"doc_key\": 44274, \"source_url\": \"\", \"split\": \"train\", \"sentences\":
[[\"THE\", \"RESUMPTION\", \"OF\", \"THE\", \"SABOTAGE\", \"AGAINST\", \"THE\",
\"ENERGY\", \"SYSTEM\", \"IS\", \"ACCOMPANIED\", \"BY\", \"THE\", \"RESUMPTION\",
\"OF\", \"REBEL\", \"ATTACKS\", \"AGAINST\", \"BROTHELS\", \",\", \"TWO\", \"OF\",
\"WHICH\", \"WERE\", \"DESTROYED\", \"IN\", \"THE\", \"LAST\", \"24\", \"HOURS\",
\",\", \"ALTHOUGH\", \"NO\", \"CASUALTIES\", \"HAVE\", \"BEEN\", \"REPORTED\", \".\"],
[\"THE\", \"BROTHELS\", \"REPORTEDLY\", \"HAVE\", \"BEEN\", \"ATTACKED\", \"BECAUSE\",
\"IT\", \"IS\", \"BELIEVED\", \"THAT\", \"WOMEN\", \"ARE\", \"EXPLOITED\", \"THERE\",
\".\"], [\"IN\", \"ADDITION\", \",\", \"URBAN\", \"GUERRILLAS\", \"BLEW\", \"UP\",
\"A\", \"PUBLIC\", \"BUS\", \"IN\", \"A\", \"NEIGHBORHOOD\", \"LOCATED\", \"NORTH\",
\"OF\", \"SAN\", \"SALVADOR\", \";\", \"NO\", \"CASUALTIES\", \"HAVE\", \"BEEN\",
\"REPORTED\", \".\"]], \"ent_spans\": [[57, 58, [\"evt007arg01attacker\", 1.0]], [62,
63, [\"evt007arg02target\", 1.0]]], \"evt_triggers\": [[59, 60,

[[\"conflict.attack.bombing\", 1.0]]], \"gold_evt_links\": [[[59, 60], [57, 58], \"evt007arg01attacker\"], [[59, 60], [62, 63], \"evt007arg02target\"]]}"

"{\"doc_key\": 44275, \"source_url\": \"\", \"split\": \"train\", \"sentences\": [[\"THE\", \"SHINING\", \"PATH\", \"GUERRILLAS\", \"ARRIVED\", \"IN\", \"HOSTIGAL\", \"LAST\", \"MONDAY\", \"MORNING\", \"AND\", \"FORCED\", \"PARISH\", \"PRIEST\", \"LUIS\", \"FUENTES\", \"TO\", \"SOUND\", \"THE\", \"ALARM\", \"IN\", \"ORDER\", \"TO\", \"GATHER\", \"THE\", \"PEOPLE\", \".\"], [\"MAYOR\", \"ARMANSIO\", \"QUISPE\", \"AND\", \"THE\", \"TWO\", \"PEASANTS\", \"WERE\", \"BROUGHT\", \"BEFORE\", \"A\", \"\\\"\", \"KANGAROO\", \"COURT\", \"\\\"\", \"BY\", \"THE\", \"SHINING\", \"PATH\", \",\", \"AND\", \"THEN\", \"KILLED\", \"BY\", \"A\", \"FIRING\", \"SQUAD\", \".\"], [\"ACCORDING\", \"TO\", \"THE\", \"LATEST\", \"CALCULATIONS\", \"MADE\", \"BY\", \"THE\", \"PERUVIAN\", \"CONGRESS\", \"PACIFICATION\", \"COMMITTEE\", \",\", \"ABOUT\", \"13,000\", \"PEOPLE\", \"HAVE\", \"BEEN\", \"KILLED\", \"IN\", \"PERU\", \"SINCE\", \"1980\", \"AS\", \"A\", \"RESULT\", \"OF\", \"POLITICAL\", \"VIOLENCE\", \".\"]], \"ent_spans\": [[3, 3, [\"evt004arg01attacker\", 1.0]], [28, 29, [\"evt004arg02target\", 1.0]]], \"evt_triggers\": [[73, 73, [[\"conflict.attack.n/a\", 1.0]]]], \"gold_evt_links\": [[[73, 73], [3, 3], \"evt004arg01attacker\"], [[73, 73], [28, 29], \"evt004arg02target\"]]}"

Non Consolidation:

My ontology handling is slightly different to Katie's because I assign victim and target to be the same due to AIDA ontology CSV thing

OUTPUT from UIUC:
Unconsolidated

1 {"doc_key": 41502, "predicted": " <arg> attacked <arg> using <arg> at <arg> place", "gold": " <arg> attacked <arg> using <arg> at EL SALVADOR place"}
2 {"doc_key": 41503, "predicted": " <arg> attacked NEARBY STRUCTURES using <arg> at <arg> place", "gold": " FPMR attacked EMBASSY BUILDING using <arg> at <arg> place"}
3 {"doc_key": 41504, "predicted": " AGUILAR MARROQUIN attacked PUBLIC BUS using <arg> at <arg> place", "gold": " <arg> attacked URBAN PUBLIC BUS using <arg> at <arg> place"}
4 {"doc_key": 41506, "predicted": " <arg> attacked <arg> using <arg> at <arg> place", "gold": " ERP attacked <arg> using HAND GRENADES at <arg> place"}
5 {"doc_key": 41507, "predicted": " <arg> attacked U.S. EMBASSY using <arg> at <arg> place", "gold": " UNIDENTIFIED PERSON attacked EMBASSY using <arg> at <arg> place"}
6 {"doc_key": 41510, "predicted": " TERRORISTS attacked TRUCK FULL OF POLICEMEN using CAR BOMB at <arg> place", "gold": " TERRORISTS ON THE DRUG CARTEL 'S PAYROLL attacked TRUCK using CAR BOMB at <arg> place"}
7 {"doc_key": 41511, "predicted": " ADOLFO MESA MENESES attacked <arg> using <arg> at COLOMBIAN place", "gold": " ADOLFO MESA MENESES attacked BERNARDO JARAMILLO OSSA using <arg> at <arg> place"}
8 {"doc_key": 41512, "predicted": " <arg> attacked SAN SALVADOR using <arg> at <arg> place", "gold": " <arg> attacked <arg> using <arg> at <arg> place"}
9 {"doc_key": 41514, "predicted": " <arg> attacked NEWSPAPER using <arg> at COLOMBIA place", "gold": " <arg> attacked NEWSPAPER using <arg> at <arg> place"}
0 {"doc_key": 41515, "predicted": " <arg> attacked COLONEL RENE EMILIO PONCE using <arg> at <arg> place", "gold": " TWO LIEUTENANTS attacked <arg> using <arg> at <arg> place"}
1 {"doc_key": 41516, "predicted": " <arg> attacked JECAR NEGHME using <arg> at <arg> place", "gold": " <arg> attacked JECAR NEGHME using <arg> at <arg> place"}
2 {"doc_key": 41517, "predicted": " ATTACKERS attacked LUIS ANTONIO MENESES BENAVIDES using <arg> at COLOMBIA place", "gold": " ATTACKERS attacked VESSEL using <arg> at <arg> place"}
3 {"doc_key": 41518, "predicted": " <arg> attacked MR. TENEDA using <arg> at <arg> place", "gold": " ATTACKERS attacked <arg> using <arg> at <arg> place"}
4 {"doc_key": 41520, "predicted": " MEDELLIN CARTEL attacked <arg> using <arg> at COLOMBIA place", "gold": " MEDELLIN CARTEL attacked <arg> using <arg> at <arg> place"}
5 {"doc_key": 41523, "predicted": " <arg> attacked <arg> using <arg> at <arg> place", "gold": " ARMED FORCES attacked <arg> using <arg> at <arg> place"}
6 {"doc_key": 41525, "predicted": " <arg> attacked <arg> using <arg> at <arg> place", "gold": " <arg> attacked PLANE using <arg> at <arg> place"}
7 {"doc_key": 41526, "predicted": " REBELS attacked ELECTRICITY POLES using <arg> at <arg> place", "gold": " REBELS attacked ELECTRICITY POLES using DYNAMITE at <arg> place"}
8 {"doc_key": 41527, "predicted": " REBELS attacked ELECTRICITY POLES using <arg> at <arg> place", "gold": " REBELS attacked ELECTRICITY POLES using DYNAMITE at <arg> place"}
9 {"doc_key": 41528, "predicted": " REBELS attacked ELECTRICITY POLES using <arg> at <arg> place", "gold": " REBELS attacked ELECTRICITY POLES using DYNAMITE at <arg> place"}
0 {"doc_key": 41529, "predicted": " REBELS attacked ELECTRICITY POLES using <arg> at <arg> place", "gold": " REBELS attacked ELECTRICITY POLES using DYNAMITE at <arg> place"}
1 {"doc_key": 41530, "predicted": " <arg> attacked <arg> using <arg> at COLOMBIAN place", "gold": " NARCOTRAFFICKERS attacked <arg> using <arg> at <arg> place"}
2 {"doc_key": 41531, "predicted": " <arg> attacked <arg> using DEATH THREATS at <arg> place", "gold": " <arg> attacked ARTURO RIVERA Y DAMAS using <arg> at <arg> place"}
3 {"doc_key": 41532, "predicted": " <arg> attacked PRIESTS using <arg> at <arg> place", "gold": " <arg> attacked <arg> using <arg> at <arg> place"}
4 {"doc_key": 41533, "predicted": " <arg> attacked <arg> using <arg> at <arg> place", "gold": " FMLN attacked <arg> using <arg> at <arg> place"}
5 {"doc_key": 41534, "predicted": " <arg> attacked <arg> using <arg> at <arg> place", "gold": " FMLN attacked <arg> using <arg> at <arg> place"}
6 {"doc_key": 41535, "predicted": " <arg> attacked <arg> using <arg> at <arg> place", "gold": " UNIFORMED MEN attacked IGNACIO ELLACURIA using <arg> at <arg> place"}
7 {"doc_key": 41536, "predicted": " <arg> attacked PIZARRO using <arg> at <arg> place", "gold": " ARMY attacked <arg> using <arg> at <arg> place"}
8 {"doc_key": 41537, "predicted": " <arg> attacked PIZARRO using <arg> at <arg> place", "gold": " SOME ARMY attacked <arg> using <arg> at <arg> place"}
9 {"doc_key": 41538, "predicted": " FMLN attacked <arg> using <arg> at <arg> place", "gold": " FMLN attacked <arg> using <arg> at <arg> place"}
0 {"doc_key": 41539, "predicted": " FMLN attacked <arg> using <arg> at <arg> place", "gold": " FMLN attacked <arg> using <arg> at <arg> place"}
1 {"doc_key": 41541, "predicted": " <arg> attacked TEACHERS using <arg> at <arg> place", "gold": " NATIONALIST REPUBLICAN ALLIANCE [ARENA] GOVERNMENT attacked CENTRAL AMERICAN UNIVERSITY using BOMB at <arg> place"}
2 {"doc_key": 41544, "predicted": " <arg> attacked <arg> using <arg> at <arg> place", "gold": " HONDURAN AND SALVADORAN attacked <arg> using <arg> at <arg> place"}
3 {"doc_key": 41545, "predicted": " <arg> attacked HONDURAN TROOPS using ROCKETS at <arg> place", "gold": " <arg> attacked HELICOPTER using MACHINEGUNS at <arg> place"}
4

Consolidated:

{"doc_key": 44403, "predicted": " GUATEMALAN GUERRILLAS attacked CENTRAL AMERICAN UNIVERSITY RECTOR IGNACIO ELLACUR", "gold": " UNIFORMED MEN attacked IGNACIO ELLACURIA using <arg> at <arg> place"}
{"doc_key": 44404, "predicted": " MURDERERS attacked ARTURO RIVERA Y DAMAS using <arg> at <arg> place", "gold": " <arg> attacked <arg> using <arg> at <arg> place"}
{"doc_key": 44405, "predicted": " MURDERERS attacked ARTURO RIVERA Y DAMAS using <arg> at <arg> place", "gold": " <arg> attacked ARTURO RIVERA Y DAMAS using <arg> at <arg> place"}
{"doc_key": 44406, "predicted": " <arg> attacked CARLOS PIZARRO using <arg> at COLOMBIAN VOTERS place", "gold": " DRUG TRAFFICKERS attacked CARLOS PIZARRO using <arg> at <arg> place"}
{"doc_key": 44407, "predicted": " REBELS attacked <arg> using RIFLE at <arg> place", "gold": " REBELS attacked ELECTRICITY POLES using DYNAMITE at <arg> place"}
{"doc_key": 44408, "predicted": " <arg> attacked <arg> using <arg> at <arg> place", "gold": " COLOMBIAN DRUG KINGPINS attacked PLANE using <arg> at <arg> place"}
{"doc_key": 44409, "predicted": " FMLN attacked COLONEL ROBERTO ARMANDO RIVERA ESCOBAR using <arg> at <arg> place", "gold": " FMLN attacked ROBERTO ARMANDO RIVERA ESCOBAR using MACHINEGUN at <arg> place"}
{"doc_key": 44410, "predicted": " COLONEL RICARDO ALFONSO CASANOVA SANDOVAL attacked JESUIT PRIESTS using", "gold": " COLONEL RICARDO ALFONSO CASANOVA SANDOVAL attacked <arg> using <arg> at <arg> place"}
{"doc_key": 44411, "predicted": " MESA MENESES attacked BERNARDO JARAMILLO OSSA using <arg> at COLOMBIAN BEACH place", "gold": " ADOLFO MESA MENESES attacked BERNARDO JARAMILLO OSSA using <arg> at <arg> place"}
{"doc_key": 44414, "predicted": " MARROQUIN attacked PUBLIC BUS using <arg> at <arg> place", "gold": " <arg> attacked URBAN PUBLIC BUS using <arg> at <arg> place"}
{"doc_key": 47240, "predicted": " MANUEL RODRIGUEZ PATRIOTIC FRONT attacked SALVADORAN EMBASSY BUILD", "gold": " FPMR attacked EMBASSY BUILDING using <arg> at <arg> place"}
{"doc_key": 47241, "predicted": " ARMED FORCES attacked <arg> using <arg> at <arg> place", "gold": " <arg> attacked <arg> using <arg> at <arg> place"}
{"doc_key": 47243, "predicted": " <arg> attacked <arg> using <arg> at COLOMBIA place", "gold": " <arg> attacked NEWSPAPER using <arg> at <arg> place"}
{"doc_key": 47244, "predicted": " <arg> attacked <arg> using <arg> at COLOMBIA place", "gold": " ATTACKERS attacked <arg> using <arg> at <arg> place"}
{"doc_key": 47246, "predicted": " MEDELLIN CARTEL attacked <arg> using <arg> at COLOMBIA place", "gold": " MEDELLIN CARTEL attacked <arg> using <arg> at <arg> place"}
{"doc_key": 47248, "predicted": " <arg> attacked FEDERAL POLICE BUILDING using TRUCK - BOMB at <arg> place", "gold": " <arg> attacked FEDERAL POLICE BUILDING using TRUCK - BOMB at <arg> place"}
{"doc_key": 47249, "predicted": " REBELS attacked <arg> using RIFLE at <arg> place", "gold": " REBELS attacked ELECTRICITY POLES using DYNAMITE at <arg> place"}
{"doc_key": 47250, "predicted": " FMLN attacked <arg> using <arg> at <arg> place", "gold": " FMLN attacked <arg> using <arg> at <arg> place"}
{"doc_key": 47251, "predicted": " FMLN attacked <arg> using <arg> at <arg> place", "gold": " FMLN attacked <arg> using <arg> at <arg> place"}

Show other outputs in html and loss graph too

How the algo works

1. Two keys, "doc_key" and "source_url," are added to the dictionary.
"doc_key" is set to the "id" field from the data at index ind.
"source_url" is initialized as an empty string.

2. The code extracts a message ID from the data, specifically from the "template" and "MESSAGE: ID" fields.Depending on the message ID prefix, the "split" key in the dictionary is set to "train," "dev," or "test."If the message ID does not match any of the expected prefixes, an exception is raised.

3.

4. The text data is extracted from the "data" field and tokenized into an array of sentences using a tokenizer

5. **HANDLE TEMPLATES WITHOUT TRIGGER:**
   If the "INCIDENT: TYPE" field in the template is equal to "*" or " *," the function returns, effectively skipping this data point.

"{\"doc_key\": 44274, \"source_url\": \"\", \"split\": \"train\", \"sentences\": [[\"THE\", \"RESUMPTION\", \"OF\", \"THE\", \"SABOTAGE\", \"AGAINST\", \"THE\", \"ENERGY\", \"SYSTEM\", \"IS\", \"ACCOMPANIED\", \"BY\", \"THE\", \"RESUMPTION\", \"OF\", \"REBEL\", \"ATTACKS\", \"AGAINST\", \"BROTHELS\", \",\", \"TWO\", \"OF\", \"WHICH\", \"WERE\", \"DESTROYED\", \"IN\", \"THE\", \"LAST\", \"24\", \"HOURS\", \",\", \"ALTHOUGH\", \"NO\", \"CASUALTIES\", \"HAVE\", \"BEEN\", \"REPORTED\", \".\"], [\"THE\", \"BROTHELS\", \"REPORTEDLY\", \"HAVE\", \"BEEN\", \"ATTACKED\", \"BECAUSE\", \"IT\", \"IS\", \"BELIEVED\", \"THAT\", \"WOMEN\", \"ARE\", \"EXPLOITED\", \"THERE\", \".\"], [\"IN\", \"ADDITION\", \",\", \"URBAN\", \"GUERRILLAS\", \"BLEW\", \"UP\", \"A\", \"PUBLIC\", \"BUS\", \"IN\", \"A\", \"NEIGHBORHOOD\", \"LOCATED\", \"NORTH\", \"OF\", \"SAN\", \"SALVADOR\", \";\", \"NO\", \"CASUALTIES\", \"HAVE\", \"BEEN\", \"REPORTED\", \".\"]], \"ent_spans\": [[57, 58, [\"evt007arg01attacker\", 1.0]], [62, 63, [\"evt007arg02target\", 1.0]]], \"evt_triggers\": [[59, 60, [[\"conflict.attack.bombing\", 1.0]]]], \"gold_evt_links\": [[[59, 60], [57, 58], \"evt007arg01attacker\"], [[59, 60], [62, 63], \"evt007arg02target\"]]}"

"{\"doc_key\": 44300, \"source_url\": \"\", \"split\": \"train\", \"sentences\": [[\"THE\", \"VICTIM\", \"WAS\", \"TAKEN\", \"TO\", \"THE\", \"AIR\", \"FORCE\", \"HOSPITAL\", \"WHERE\", \"HE\", \"UNFORTUNATELY\", \"PASSED\", \"AWAY\", \".\"], [\"3\", \".\"], [\"THE\", \"AUTHORITIES\", \"ARE\", \"INVESTIGATING\", \"THE\", \"ATTACK\", \",\", \"AND\", \"THEY\", \"ARE\",

\"CONDUCTING\", \"THE\", \"APPROPRIATE\", \"OPERATIONS\", \"TO\", \"CAPTURE\", \"THE\", \"CRIMINALS\", \".\"], [\"[\", \"DATED\", \"AND\", \"SIGNED\", \"]\", \"LIMA\", \",\", \"9\", \"JANUARY\", \"1990\", \".\"], [\"DEFENSE\", \"MINISTRY\", \"COMMUNICATIONS\", \"OFFICE\", \".\"]], \"ent_spans\": [], \"evt_triggers\": [[22, 22, [[\"conflict.attack.n/a\", 1.0]]]], \"gold_evt_links\": []}"

"{\"doc_key\": 44301, \"source_url\": \"\", \"split\": \"train\", \"sentences\": [[\"SANDOVAL\", \"POINTED\", \"OUT\", \"THAT\", \"\\\"\", \"MORE\", \"THAN\", \"SEVEN\", \"PEOPLE\", \"WERE\", \"INVOLVED\", \"IN\", \"THE\", \"MURDER\", \",\", \"AND\", \"TWO\", \"PERPETRATED\", \"IT\", \".\", \"\\\"\"], [\"HE\", \"ADDED\", \"THAT\", \"TWO\", \"DIFFERENT\", \"GROUPS\", \"ACTED\", \",\", \"ONE\", \"THAT\", \"MURDERED\", \"THE\", \"CLERGYMEN\", \"AND\", \"ANOTHER\", \"THAT\", \"MURDERED\", \"TWO\", \"MAIDS\", \",\", \"ACCORDING\", \"TO\", \"PRELIMINARY\", \"RESULTS\", \"ISSUED\", \"BY\", \"THE\", \"CRIME\", \"COMMISSION\", \"IN\", \"CHARGE\", \"OF\", \"SOLVING\", \"THE\", \"CASE\", \".\"], [\"\\\"\", \"THERE\", \"WAS\", \"NO\", \"TORTURE\", \"OR\", \"MISTREATMENT\", \",\", \"BUT\", \"WORDS\", \"WERE\", \"EXCHANGED\", \"\\\"\", \"BETWEEN\", \"THE\", \"MURDERERS\", \"AND\", \"VICTIMS\", \",\", \"THE\", \"SPOKESMAN\", \"SAID\", \",\", \"ADDING\", \"THAT\", \"FIVE\", \"OF\", \"THE\", \"MURDERED\", \"JESUITS\", \"HAD\", \"RECEIVED\", \"\\\"\", \"FINISHING\", \"SHOTS\", \",\", \"\\\"\", \"ACCORDING\", \"TO\", \"AUTOPSIES\", \"CARRIED\", \"OUT\", \"BY\", \"SALVADORAN\", \"FORENSIC\", \"DOCTORS\", \".\"], [\"ACCORDING\", \"TO\", \"TESTIMONY\", \"THE\", \"COMMISSION\", \"GATHERED\", \"FROM\", \"115\", \"MEMBERS\", \"OF\", \"THE\", \"ARMY\", \"AND\", \"103\", \"CIVILIANS\", \",\", \"THE\", \"MURDER\", \"TOOK\", \"PLACE\", \"BETWEEN\", \"0200\", \"AND\", \"0300\", \",\", \"AND\", \"THERE\", \"WERE\", \"\\\"\", \"EXPLOSIONS\", \",\", \"MACHINE\", \"-\", \"GUN\", \"BLASTS\", \",\", \"AND\", \"SHOTS\", \",\", \"\\\"\", \"SANDOVAL\", \"SAID\", \".\"], [\"THE\", \"SPOKESMAN\", \"ADDED\", \"THAT\", \"2\", \"OFFICERS\", \"AND\", \"45\", \"SOLDIERS\", \"HAVE\", \"BEEN\", \"\\\"\", \"CONFINED\", \"IN\", \"VARIOUS\", \"HEADQUARTERS\", \"PENDING\", \"INVESTIGATION\", \".\", \"\\\"\"]], \"ent_spans\": [[135, 137, [\"evt004arg03instrument\", 1.0]], [5, 8, [\"evt004arg01attacker\", 1.0]]], \"evt_triggers\": [[85, 85, [[\"conflict.attack.n/a\", 1.0]]]], \"gold_evt_links\": [[[85, 85], [135, 137], \"evt004arg03instrument\"], [[85, 85], [5, 8], \"evt004arg01attacker\"]]}"

1. Do we have whole document not just some sentences?
2. Come up with something for no triggers
3. 44278\
4. 5 sentences limit for RAMS
5.

Improvements