Thread Labelling

April 16, 2022

1 Setup

```
[2]: import pandas as pd
```

```
[3]: import pickle as pkl
```

```
[30]: import numpy as np
```

```
[20]: from tqdm import tqdm
```

2 Load Data

```
[4]: inverted = pkl.load(open('data/inverted_dict.pkl','rb'))
```

```
[5]: type(inverted)
```

```
[5]: dict
```

```
[6]: inverted
```

```
[6]: {'aatrox': 'aatrox',
     'atrox': 'aatrox',
     'ari': 'ahri',
     'ahri': 'ahri',
     'akali': 'akali',
     'ali': 'alistar',
     'alistar': 'alistar',
     'amumu': 'amumu',
     'anivia': 'anivia',
     'annie': 'annie',
     'aphelios': 'aphelios',
     'aphel': 'aphelios',
     'ashe': 'ashe',
     'asol': 'aurelion sol',
     'aurelion sol': 'aurelion sol',
     'sol': 'aurelion sol',
     'aurelion': 'aurelion sol',
```

```
'azir': 'azir',
'bard': 'bard',
'blitz': 'blitzcrank',
'blitzcrank': 'blitzcrank',
'brand': 'brand',
'braum': 'braum',
'caitlyn': 'caitlyn',
'cait': 'caitlyn',
'camille': 'camille',
'cassiopeia': 'cassiopeia',
'cass': 'cassiopeia',
"cho'gath": "cho'gath",
'chogath': "cho'gath",
'cho': "cho'gath",
'gath': "cho'gath",
'corki': 'corki',
'darius': 'darius',
'diana': 'diana',
'mundo': 'dr. mundo',
'dr. mundo': 'dr. mundo',
'dr mundo': 'dr. mundo',
'draven': 'draven',
'ekko': 'ekko',
'elise': 'elise',
'eve': 'evelynn',
'evelynn': 'evelynn',
'ez': 'ezreal',
'ezreal': 'ezreal',
'fiddlesticks': 'fiddlesticks',
'fid': 'fiddlesticks',
'fiddle': 'fiddlesticks',
'fiora': 'fiora',
'fizz': 'fizz',
'galio': 'galio',
'gangplank': 'gangplank',
'gp': 'gangplank',
'garen': 'garen',
'gnar': 'gnar',
'gragas': 'gragas',
'grag': 'gragas',
'graves': 'graves',
'hec': 'hecarim',
'hecarim': 'hecarim',
'heimerdinger': 'heimerdinger',
'heim': 'heimerdinger',
'illaoi': 'illaoi',
'irel': 'irelia',
```

```
'irelia': 'irelia',
'ivern': 'ivern',
'janna': 'janna',
'jarvan iv': 'jarvan iv',
'jarvan': 'jarvan iv',
'j4': 'jarvan iv',
'jax': 'jax',
'jayce': 'jayce',
'jhin': 'jhin',
'powder': 'jinx',
'jinx': 'jinx',
'sa': "kai'sa",
"kai'sa": "kai'sa",
'kaisa': "kai'sa",
'kai': "kai'sa",
'kalista': 'kalista',
'kali': 'kalista',
'karma': 'karma',
'karthus': 'karthus',
'karth': 'karthus',
'kass': 'kassadin',
'kassadin': 'kassadin',
'kat': 'katarina',
'katarina': 'katarina',
'kayle': 'kayle',
'kayn': 'kayn',
'kennen': 'kennen',
'kha': "kha'zix",
'k6': "kha'zix",
'zix': "kha'zix",
"kha'zix": "kha'zix",
'khazix': "kha'zix",
'kindred': 'kindred',
'kled': 'kled',
'maw': "kog'maw",
'kog': "kog'maw",
'kogmaw': "kog'maw",
"kog'maw": "kog'maw",
'lb': 'leblanc',
'leblanc': 'leblanc',
'leesin': 'lee sin',
'lee': 'lee sin',
'sin': 'lee sin',
'lee sin': 'lee sin',
'leona': 'leona',
'lillia': 'lillia',
'lissandra': 'lissandra',
```

```
'liss': 'lissandra',
'lucian': 'lucian',
'lulu': 'lulu',
'lux': 'lux',
'malphite': 'malphite',
'malph': 'malphite',
'malzahar': 'malzahar',
'malz': 'malzahar',
'maokai': 'maokai',
'mao': 'maokai',
'masteryi': 'master yi',
'yi': 'master yi',
'master yi': 'master yi',
'master': 'master yi',
'mf': 'miss fortune',
'fortune': 'miss fortune',
'miss': 'miss fortune',
'missfortune': 'miss fortune',
'miss fortune': 'miss fortune',
'morde': 'mordekaiser',
'mord': 'mordekaiser',
'mordekaiser': 'mordekaiser',
'morgana': 'morgana',
'morg': 'morgana',
'nami': 'nami',
'nasus': 'nasus',
'naut': 'nautilus',
'nautilus': 'nautilus',
'neeko': 'neeko',
'nidalee': 'nidalee',
'nid': 'nidalee',
'noc': 'nocturne',
'nocturne': 'nocturne',
'nunu': 'nunu and willump',
'nunu and willump': 'nunu and willump',
'willump': 'nunu and willump',
'olaf': 'olaf',
'ori': 'orianna',
'orianna': 'orianna',
'ornn': 'ornn',
'panth': 'pantheon',
'pantheon': 'pantheon',
'poppy': 'poppy',
'pyke': 'pyke',
'qiyana': 'qiyana',
'quinn': 'quinn',
'rakan': 'rakan',
```

```
'rammus': 'rammus',
'rek': "rek'sai",
'reksai': "rek'sai",
'sai': "rek'sai",
"rek'sai": "rek'sai",
'rell': 'rell',
'renekton': 'renekton',
'renek': 'renekton',
'rengar': 'rengar',
'riven': 'riven',
'riv': 'riven',
'rumble': 'rumble',
'ryze': 'ryze',
'sam': 'samira',
'samira': 'samira',
'sejuani': 'sejuani',
'sej': 'sejuani',
'senna': 'senna',
'seraphine': 'seraphine',
'sera': 'seraphine',
'sett': 'sett',
'shaco': 'shaco',
'shen': 'shen',
'shyvana': 'shyvana',
'shyv': 'shyvana',
'singed': 'singed',
'sion': 'sion',
'sivir': 'sivir',
'skarner': 'skarner',
'sona': 'sona',
'raka': 'soraka',
'soraka': 'soraka',
'swain': 'swain',
'sylas': 'sylas',
'syndra': 'syndra',
'tahm kench': 'tahm kench',
'tahm': 'tahm kench',
'kench': 'tahm kench',
'tahmkench': 'tahm kench',
'taliyah': 'taliyah',
'talon': 'talon',
'taric': 'taric',
'teemo': 'teemo',
'thresh': 'thresh',
'trist': 'tristana',
'tristana': 'tristana',
'trund': 'trundle',
```

```
'trundle': 'trundle',
'tryndamere': 'tryndamere',
'trynd': 'tryndamere',
'tf': 'twisted fate',
'twisted': 'twisted fate',
'twistedfate': 'twisted fate',
'fate': 'twisted fate',
'twisted fate': 'twisted fate',
'twitch': 'twitch',
'udyr': 'udyr',
'urgot': 'urgot',
'varus': 'varus',
'vayne': 'vayne',
'veigar': 'veigar',
'veig': 'veigar',
'vel': "vel'koz",
"vel'koz": "vel'koz",
'koz': "vel'koz",
'velkoz': "vel'koz",
'vi': 'vi',
'vy': 'vi',
'vik': 'viktor',
'viktor': 'viktor',
'vladimir': 'vladimir',
'vlad': 'vladimir',
'voli': 'volibear',
'volibear': 'volibear',
'ww': 'warwick',
'warwick': 'warwick',
'wuk': 'wukong',
'wukong': 'wukong',
'xayah': 'xayah',
'xerath': 'xerath',
'xer': 'xerath',
'xin zhao': 'xin zhao',
'zhao': 'xin zhao',
'xinzhao': 'xin zhao',
'xin': 'xin zhao',
'yas': 'yasuo',
'yasuo': 'yasuo',
'yone': 'yone',
'yorick': 'yorick',
'yuumi': 'yuumi',
'zac': 'zac',
'zed': 'zed',
'ziggs': 'ziggs',
'zilean': 'zilean',
```

```
    'zil': 'zilean',
    'zoe': 'zoe',
    'zyra': 'zyra',
    'renata glasc': 'renata glasc',
    'glasc': 'renata glasc',
    'renataglasc': 'renata glasc',
    'renata': 'renata glasc',
    'zeri': 'zeri',
    'vex': 'vex',
    'akshan': 'akshan',
    'gwen': 'gwen',
    'viego': 'viego'}
```

[7]: `df_comments = pd.read_csv('data/comments_dataframe.csv', index_col=[0])`

[8]: `df_comments`

[8]:
```
        submission_id comment_id  \
0              o00m44     h1si8vd
1              o00m44     h1sil8e
2              o00m44     h1sigkk
3              o00m44     h1sttyw
4              o00m44     h1sqn3p
...               ...         ...
819317         t3tx4z     hyuijui
819318         t3tx4z     hyujigw
819319         t3tx4z     hyujjc3
819320         t3twyu     hyujgh3
819321         t3twhr     hyuikvu


                                             comment_text
0          Roam by having the wave pushed to enemy turret…
1          The first step to roaming is having good wave …
2          Try not to roam when you think you won't get a…
3                                    Roam less. farm more
4          1 roam kill is one missed wave. One missed wav…
...                                                    ...
819317     Look in event viewer in windows to see what is…
819318     Hi /u/-CrestiaBell. Thank you for participatin…
819319     Im on a similar laptop with W11 and no problem…
819320     Hi /u/DiabloDJ. Thank you for participating in…
819321     Hi /u/NotARussian421. Thank you for participat…

[819322 rows x 3 columns]
```

[9]: `df_submissions = pd.read_csv('data/submissions_dataframe.csv',index_col=[0])`

```
[10]: df_submissions

[10]:         created_utc       id  \
      0        1623715013   o00m44
      1        1623714840   o00k4w
      2        1623714789   o00jk8
      3        1623714743   o00j2e
      4        1623714562   o00gyo
      …                …        …
      143562   1646093044   t3txtq
      143563   1646092994   t3tx4z
      143564   1646092983   t3twyu
      143565   1646092952   t3twhr
      143566   1646092817   t3tuqh


                                                       title  \
      0                                    I dont cs when im fed
      1             Totally planned prediction 2021 outplay that i…
      2         UCAM Esports vs MAD Lions Madrid - LEAGUE OF L…
      3                          One hella lucky play on my le blanc
      4                             A digital painting of Wolf!!
      …                                                       …
      143562                                      The Double Steal
      143563                         Frequent crashes on Windows 11.
      143564   Tarzaned leaving game after dying lvl 1, flame…
      143565   Do not criticize russian aggression or Vladimi…
      143566                                      Lunar Revel Pass


                                                 url       created  \
      0        https://www.reddit.com/r/leagueoflegends/comme…  1.623740e+09
      1                      https://v.redd.it/iqthzskakb571  1.623740e+09
      2        https://youtube.com/watch?v=wfHM4nX2z-A&amp;fe…  1.623740e+09
      3                      https://v.redd.it/c9puqrqyjb571  1.623740e+09
      4                  https://i.redd.it/73ln16dhjb571.jpg  1.623740e+09
      …                                                    …         …
      143562                https://v.redd.it/f11af9f4ynk81  1.646118e+09
      143563   https://www.reddit.com/r/leagueoflegends/comme…  1.646118e+09
      143564                https://v.redd.it/tpem6nm0ynk81  1.646118e+09
      143565   https://www.reddit.com/r/leagueoflegends/comme…  1.646118e+09
      143566   https://www.reddit.com/r/leagueoflegends/comme…  1.646118e+09


                                                 d_              datetime
      0        {'created_utc': 1623715013, 'id': 'o00m44', 't…  2021-06-14 23:56:53
      1        {'created_utc': 1623714840, 'id': 'o00k4w', 't…  2021-06-14 23:54:00
      2        {'created_utc': 1623714789, 'id': 'o00jk8', 't…  2021-06-14 23:53:09
      3        {'created_utc': 1623714743, 'id': 'o00j2e', 't…  2021-06-14 23:52:23
      4        {'created_utc': 1623714562, 'id': 'o00gyo', 't…  2021-06-14 23:49:22
```

```
…                                          …                        …
143562  {'created_utc': 1646093044, 'id': 't3txtq', 't…  2022-03-01 00:04:04
143563  {'created_utc': 1646092994, 'id': 't3tx4z', 't…  2022-03-01 00:03:14
143564  {'created_utc': 1646092983, 'id': 't3twyu', 't…  2022-03-01 00:03:03
143565  {'created_utc': 1646092952, 'id': 't3twhr', 't…  2022-03-01 00:02:32
143566  {'created_utc': 1646092817, 'id': 't3tuqh', 't…  2022-03-01 00:00:17

[143567 rows x 7 columns]
```

3 Wrangling

```
[11]: df_processing = df_comments.drop(columns=['comment_id'])
```

```
[12]: df_processing
```

```
[12]:        submission_id                                     comment_text
       0             o00m44  Roam by having the wave pushed to enemy turret…
       1             o00m44  The first step to roaming is having good wave …
       2             o00m44  Try not to roam when you think you won't get a…
       3             o00m44                                Roam less. farm more
       4             o00m44  1 roam kill is one missed wave. One missed wav…
       …                …                                                  …
       819317        t3tx4z  Look in event viewer in windows to see what is…
       819318        t3tx4z  Hi /u/-CrestiaBell. Thank you for participatin…
       819319        t3tx4z  Im on a similar laptop with W11 and no problem…
       819320        t3twyu  Hi /u/DiabloDJ. Thank you for participating in…
       819321        t3twhr  Hi /u/NotARussian421. Thank you for participat…

[819322 rows x 2 columns]
```

```
[13]: df_processing['comment_text'] = df_comments.
      ↪groupby('submission_id')['comment_text'].transform(lambda x: ' '.join(x))
      df_processing = df_processing[['submission_id','comment_text']].
      ↪drop_duplicates()
```

```
[14]: df_processing
```

```
[14]:        submission_id                                     comment_text
       0             o00m44  Roam by having the wave pushed to enemy turret…
       10            o00j2e  That Xerath is fucking lost LMFAO. that's not …
       22            o00gbv  Your thread has a disallowed title structure, …
       23            o00c4g  I think it's just bad luck man unless they are…
       27            o00azn  Remove feast\n\nRemove Olaf ult\n\nRemove tryn…
       …                …                                                  …
       819304        t3u808  It's so interesting how low elo is so spiteful…
       819308        t3txtq  i wish people would just pull drags out especi…
```

9

```
819317        t3tx4z  Look in event viewer in windows to see what is…
819320        t3twyu  Hi /u/DiabloDJ. Thank you for participating in…
819321        t3twhr  Hi /u/NotARussian421. Thank you for participat…

[87728 rows x 2 columns]
```

[15]: ```python
df_processing = df_processing.reset_index(drop=True)
```

[16]: ```python
df_processing
```

[16]:
```
       submission_id                                   comment_text
0             o00m44  Roam by having the wave pushed to enemy turret…
1             o00j2e  That Xerath is fucking lost LMFAO. that's not …
2             o00gbv  Your thread has a disallowed title structure, …
3             o00c4g  I think it's just bad luck man unless they are…
4             o00azn  Remove feast\n\nRemove Olaf ult\n\nRemove tryn…
...              ...                                            ...
87723         t3u808  It's so interesting how low elo is so spiteful…
87724         t3txtq  i wish people would just pull drags out especi…
87725         t3tx4z  Look in event viewer in windows to see what is…
87726         t3twyu  Hi /u/DiabloDJ. Thank you for participating in…
87727         t3twhr  Hi /u/NotARussian421. Thank you for participat…

[87728 rows x 2 columns]
```

4 Labelling

[17]: ```python
champions_list =␣
↪"Aatrox,Ahri,Akali,Alistar,Amumu,Anivia,Annie,Aphelios,Ashe,Aurelion␣
↪Sol,Azir,Bard,Blitzcrank,Brand,Braum,Caitlyn,Camille,Cassiopeia,Cho'Gath,Corki,Darius,Diana
↪␣
↪Mundo,Draven,Ekko,Elise,Evelynn,Ezreal,Fiddlesticks,Fiora,Fizz,Galio,Gangplank,Garen,Gnar,G
↪IV,Jax,Jayce,Jhin,Jinx,Kai'Sa,Kalista,Karma,Karthus,Kassadin,Katarina,Kayle,Kayn,Kennen,Kha
↪Sin,Leona,Lillia,Lissandra,Lucian,Lulu,Lux,Malphite,Malzahar,Maokai,Master␣
↪Yi,Miss␣
↪Fortune,Mordekaiser,Morgana,Nami,Nasus,Nautilus,Neeko,Nidalee,Nocturne,Nunu␣
↪and␣
↪Willump,Olaf,Orianna,Ornn,Pantheon,Poppy,Pyke,Qiyana,Quinn,Rakan,Rammus,Rek'Sai,Rell,Renekt
↪Kench,Taliyah,Talon,Taric,Teemo,Thresh,Tristana,Trundle,Tryndamere,Twisted␣
↪Fate,Twitch,Udyr,Urgot,Varus,Vayne,Veigar,Vel'Koz,Vi,Viktor,Vladimir,Volibear,Warwick,Wukon
↪Zhao,Yasuo,Yone,Yorick,Yuumi,Zac,Zed,Ziggs,Zilean,Zoe,Zyra"
champions_list = champions_list.split(",")
champions_list.append("Renata Glasc")
champions_list.append("Zeri")
champions_list.append("Vex")
champions_list.append("Akshan")
```

```
    champions_list.append("Gwen")
    champions_list.append("Viego")
    champions_list = [x.lower() for x in champions_list]
```

[19]:
```
pkl.dump(champions_list, open('champions_list.pkl','wb'))
```

[21]:
```
topic_champions = list()

for index, row in tqdm(df_processing.iterrows()):
    tokens = row['comment_text'].lower().split()

    counter_dict = {x : 0 for x in champions_list}

    for token in tokens:
        if token in inverted.keys():
            counter_dict[inverted[token]] += 1

    champions = sorted(counter_dict.items(),
                       key = lambda x: x[1],
                       reverse = True)

    if len(champions) == 0:
        topic_champions.append(None)
    elif len(champions) == 1:
        topic_champions.append([champions[0][0]])
    elif champions[0][1] == champions[1][1]:
        topic_champions.append([champions[0][0], champions[1][0]])
    elif len(champions) >= 3 and champions[0][1] == champions[1][1] and␣
↪champions[2][1]:
        topic_champions.append([champions[0][0], champions[1][0],␣
↪champions[2][1]])
    else:
        topic_champions.append([champions[0][0]])

df_processing['topic_champion'] = topic_champions
```

87728it [00:08, 10878.39it/s]

[22]:
```
df_processing
```

[22]:
```
        submission_id                                 comment_text  \
0              o00m44   Roam by having the wave pushed to enemy turret…
1              o00j2e   That Xerath is fucking lost LMFAO. that's not …
2              o00gbv   Your thread has a disallowed title structure, …
3              o00c4g   I think it's just bad luck man unless they are…
4              o00azn   Remove feast\n\nRemove Olaf ult\n\nRemove tryn…
…                  …                                              …
87723          t3u808   It's so interesting how low elo is so spiteful…
```

11

```
87724        t3txtq  i wish people would just pull drags out especi…
87725        t3tx4z  Look in event viewer in windows to see what is…
87726        t3twyu  Hi /u/DiabloDJ. Thank you for participating in…
87727        t3twhr  Hi /u/NotARussian421. Thank you for participat…

             topic_champion
0            [twisted fate]
1                  [xerath]
2            [aatrox, ahri]
3            [aatrox, ahri]
4         [olaf, tryndamere]
…                         …
87723        [aatrox, ahri]
87724        [jinx, kha'zix]
87725        [aatrox, ahri]
87726        [aatrox, ahri]
87727        [aatrox, ahri]

[87728 rows x 3 columns]
```

[23]:
```python
df_processing.to_csv('data/df_processing.csv')
```

5   Conversion to 1NF

[31]:
```python
lengths = list(map(len, df_processing['topic_champion'].values))

df_labels_1nf = pd.DataFrame(
    {'submission_id': np.repeat(df_processing['submission_id'], lengths),
     'topic_champion': np.concatenate(df_processing['topic_champion'].values)
    }
)
```

[32]:
```python
df_labels_1nf
```

[32]:
```
       submission_id topic_champion
0            o00m44    twisted fate
1            o00j2e          xerath
2            o00gbv          aatrox
2            o00gbv            ahri
3            o00c4g          aatrox
…                 …               …
87725        t3tx4z            ahri
87726        t3twyu          aatrox
87726        t3twyu            ahri
87727        t3twhr          aatrox
87727        t3twhr            ahri
```

```
[148662 rows x 2 columns]
```

[33]: `df_labels_1nf.to_csv('data/df_labels_1nf.csv')`