

World Cup 2022 in various media outlets

Adam Janczyszyn, Jakub Bandurski

University of Warsaw

jbandurski@student.uw.edu.pl

ajanczyszyn@student.uw.edu.pl

February 10, 2023

Overview

- 1 Motivation
- 2 Data
- 3 Preprocessing
- 4 LDA
 - CNN
 - Qatar
- 5 BERT
 - CNN
 - Qatar
- 6 Conclusions
- 7 References

Motivation

The main motivation behind the work was the desire to compare how the World Cup was perceived by the Qatari government media and the recognised global media, CNN.

We also wanted to check the difference in the occurrence of non-sports topics in relation to the 2022 World Cup.

Lastly, as football fans, we wanted to check and summarise the key events of the World Cup in Qatar.

We scrapped articles relating to World Cup 2022 from both CNN, qatar2022.qa and twitter. So far we analyzed first and second dataset.

The time span is respectively April 2022 – January 2023.

Overall we have 370 articles from CNN¹ and 157 articles from qatar2022.qa².

¹[cnn.com](https://www.cnn.com)

²qatar2022.qa/en/news

- Data cleaning
- Tokenization
- Stemming
- Standard stopwords with additional:
 - world, cup, fifa, his, was, were etc.
- Ngrams 1-3
- TF approach
- TF-IDF approach



Number of topics $k = 22$ for LDA was chosen with use of coherence.

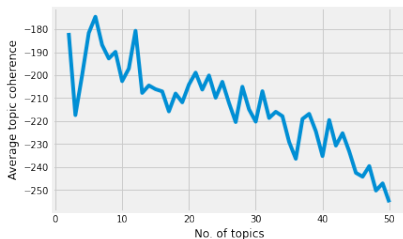


Figure: Coherence for TF

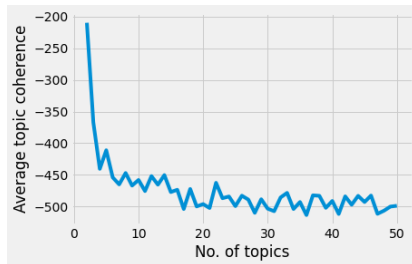


Figure: Coherence for TF-IDF

Parameters Optimization TF CNN

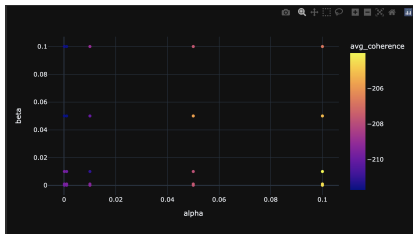


Figure: Plot of avg. coherence vs α and β

| | alpha | beta | vectorizer | avg_coherence |
|----|--------|--------|------------|---------------|
| 44 | 0.1000 | 0.0100 | tf | -204.006580 |
| 40 | 0.1000 | 0.0001 | tf | -204.455686 |
| 42 | 0.1000 | 0.0010 | tf | -204.461116 |
| 46 | 0.1000 | 0.0500 | tf | -205.221075 |
| 36 | 0.0500 | 0.0500 | tf | -205.729977 |
| 48 | 0.1000 | 0.1000 | tf | -206.775352 |
| 32 | 0.0500 | 0.0010 | tf | -207.616414 |
| 38 | 0.0500 | 0.1000 | tf | -207.643828 |
| 30 | 0.0500 | 0.0001 | tf | -207.683504 |
| 34 | 0.0500 | 0.0100 | tf | -207.803109 |
| 28 | 0.0100 | 0.1000 | tf | -208.001728 |

50 rows x 4 columns [Open in new tab](#)

Figure: Table of avg. coherence vs α and β

Conclusion $\alpha = 0.1, \beta = 0.01$.

Parameters Optimization TF-IDF CNN

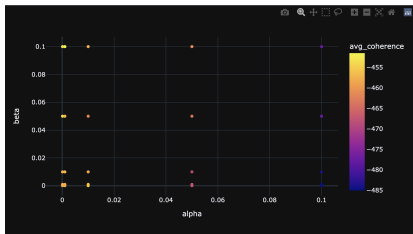


Figure: Plot of avg. coherence vs α and β

| | alpha | beta | vectorizer | avg_coherence |
|----|--------|--------|------------|---------------|
| 19 | 0.0010 | 0.1000 | tf-idf | -451.520445 |
| 9 | 0.0001 | 0.1000 | tf-idf | -453.476329 |
| 7 | 0.0001 | 0.0500 | tf-idf | -453.645025 |
| 21 | 0.0100 | 0.0001 | tf-idf | -454.477188 |
| 23 | 0.0100 | 0.0010 | tf-idf | -455.874104 |
| 17 | 0.0010 | 0.0500 | tf-idf | -456.424172 |
| 15 | 0.0010 | 0.0100 | tf-idf | -456.680062 |
| 5 | 0.0001 | 0.0100 | tf-idf | -457.332053 |
| 3 | 0.0001 | 0.0010 | tf-idf | -457.594023 |
| 29 | 0.0100 | 0.1000 | tf-idf | -457.767595 |

Figure: Table of avg. coherence vs α and β

Conclusion $\alpha = 0.001$, $\beta = 0.1$.

TF Word Clouds CNN



LDA TF Topics CNN

```
Topic 0: ['match', 'team', 'argentina', 'game', 'goal', 'first', 'messi', 'group', 'france', 'win']
Topic 1: ['soccer', 'us', 'pelé', 'team', 'before', 'al', 'players', 'saudi', 'statement', 'ronaldo']
Topic 2: ['messi', 'argentina', 'first', 'game', 'match', 'goal', 'france', 'two', 'team', 'win']
Topic 3: ['brazil', 'match', 'portugal', 'ronaldo', 'first', 'south', 'team', 'goal', 'korea', 'south korea']
Topic 4: ['workers', 'rights', 'migrant', 'migrant workers', 'human rights', 'human', 'tournament', 'fifa', 'qatars', 'qatari']
Topic 5: ['team', 'messi', 'argentina', 'france', 'match', 'first', 'tournament', 'final', 'two', 'year']
Topic 6: ['argentina', 'football', 'live', 'match', 'two', 'soccer', 'tournament', 'like', 'game', 'first']
Topic 7: ['russian', 'ukraine', 'ukrainian', 'russia', 'us', 'city', 'also', 'war', 'president', 'forces']
Topic 8: ['us', 'soccer', 'team', 'people', 'cnn', 'told', 'two', 'football', 'fans', 'also']
Topic 9: ['people', 'tournament', 'al', 'portugal', 'also', 'team', 'cnn', 'human rights', 'around', 'year']
Topic 10: ['first', 'match', 'tournament', 'argentina', 'morocco', 'team', 'football', 'two', 'year', 'france']
Topic 11: ['morocco', 'first', 'tournament', 'team', 'year', 'workers', 'goal', 'final', 'argentina', 'two']
Topic 12: ['game', 'fifa', 'tournament', 'first', 'argentina', 'soccer', 'france', 'messi', 'cnn', 'player']
Topic 13: ['first', 'football', 'year', 'team', 'match', 'also', 'messi', 'national', 'mbappé', 'tournament']
Topic 14: ['football', 'us', 'fifa', 'iran', 'people', 'team', 'soccer', 'cnn', 'fans', 'players']
Topic 15: ['wahl', 'first', 'before', 'soccer', 'match', 'grant', 'final', 'cnn', 'pelé', 'told']
Topic 16: ['team', 'first', 'argentina', 'portugal', 'morocco', 'goal', 'tournament', 'teams', 'match', 'messi']
Topic 17: ['people', 'also', 'us', 'soccer', 'morocco', 'two', 'cnn', 'team', 'match', 'told']
Topic 18: ['tournament', 'morocco', 'croatia', 'first', 'team', 'game', 'also', 'two', 'people', 'players']
Topic 19: ['argentina', 'match', 'team', 'first', 'messi', 'brazil', 'also', 'vs', 'president', 'told']
Topic 20: ['argentina', 'first', 'match', 'messi', 'team', 'croatia', 'game', 'tournament', 'spain', 'morocco']
Topic 21: ['first', 'match', 'argentina', 'messi', 'morocco', 'team', 'football', 'two', 'goal', 'tournament']
```

Topic 1:

['soccer', 'us', 'pelé', 'team', 'before', 'al', 'players', 'saudi', 'statement', 'ronaldo']

Topic 4:

['workers', 'rights', 'migrant', 'migrant workers', 'human rights', 'human', 'tournament', 'fifa', 'qatars', 'qatari']

Topic 7:

['russian', 'ukraine', 'ukrainian', 'russia', 'us', 'city', 'also', 'war', 'president', 'forces']

TF-IDF Word Clouds CNN



LDA TF-IDF Topics CNN

```
Topic 0: ['messi', 'match', 'holders', 'us_national', 'statement_said', 'argentina', 'amini', 'stadium_monday', 'least_five', 'gone_play']
Topic 1: ['milan', 'need_know', 'costa_rica', 'given', 'players', 'achieving', 'mariupol', 'program', 'dominik_livaković', 'before_first']
Topic 2: ['ronaldo', 'messi', 'because_something', 'magic', 'angel_di', 'croatias', 'new_york_times', 'final_whistle_during', 'cnn', 'rashford']
Topic 3: ['fifa', 'ronaldo', 'rights', 'human_rights', 'ten_hag', 'match', 'saudi', 'host', 'commit', 'does_get']
Topic 4: ['workers', 'migrant_workers', 'us', 'drugs', 'construction', 'migrant', 'president', 'berhalter', 'fifa', 'wednesday']
Topic 5: ['golf', 'per', 'earlier_today', 'grant', 'finals', 'heart', 'term', 'win_game', 'said_referring', 'feet']
Topic 6: ['make_match', 'licensed_venues', 'touched', 'contributed_reporting_post', 'sides_first', 'along_way', 'human_rights_death', 'impossible', 'hit_back', 'slowly']
Topic 7: ['ukraine', 'russian', 'us', 'fifa', 'team', 'ukrainian', 'cnn', 'country', 'tournament', 'fans']
Topic 8: ['messi_scored', 'two', 'punishable', 'condemned', 'bus', 'check_moving', 'brace', 'earlier_monday', 'adoring', 'see']
Topic 9: ['todays_match', 'iran', 'human_rights', 'iranian', 'iranian_football', 'cavallo', 'us_armband', 'news', 'portugal', 'shock_grief']
Topic 10: ['season', 'chain', 'said_during', 'crowd', 'reasons', 'match', 'south_korea', 'league', 'first', 'france']
Topic 11: ['argentina', 'morocco', 'jordan', 'figure', 'team', 'statement', 'kingdoms', 'tickets', 'workers', 'historic']
Topic 12: ['off_the', 'fifa', 'states_mens', 'women', 'working_conditions', 'iranian_national_team', 'irans_first', 'sergião', 'tournament', 'attacker']
Topic 13: ['national', 'training', 'counterattack', 'iran', 'pundits', 'options', 'spell', 'forum', 'could_see', 'valiantly']
Topic 14: ['team', 'match', 'us', 'messi', 'fifa', 'first', 'tournament', 'game', 'football', 'argentina']
Topic 15: ['soccer', 'us', 'match', 'russian', 'wahl', 'pelé', 'israel', 'years', 'gounder', 'before']
Topic 16: ['argentina', 'goal', 'messi', 'team', 'match', 'first', 'half', 'game', 'win', 'mbappé']
Topic 17: ['flights', 'airlines', 'fans', 'issued_statement', 'manager_fernando_santos', 'establishment', 'three_countries', 'achraf_hakimi', 'average', 'drink']
Topic 18: ['jakob_jensen', 'happy', 'danilo', 'qatar_also', 'wage', 'played', 'advocates', 'football_governing_body', 'earlier_year', 'return_home']
Topic 19: ['appearing_shadow_domestic', 'took_control', 'agency_tass', 'xherdan_shaqiri', 'vladimir_putin', 'wanted_change_official', 'evening', 'recently_told', 'lifting_trophy', 'us_soccer_federation']
Topic 20: ['information_categorically', 'spare', 'three_years', 'manager_hansi', 'football_told', 'fuel', 'flagship', 'bukayo', 'unleashed', 'moon']
Topic 21: ['uae', 'kicks', 'tchouaméni', 'resort', 'build', 'buoyed', 'efficiency', 'iran', 'real_madrid', 'forty']
```

Topic 3:

['fifa', 'ronaldo', 'rights', 'human_rights', 'ten_hag', 'match', 'saudi', 'host', 'commit', 'does_get']

Topic 4:

['workers', 'migrant_workers', 'us', 'drugs', 'construction', 'migrant', 'president', 'berhalter', 'fifa', 'wednesday']

Topic 6:

['make_match', 'licensed_venues', 'touched', 'contributed_reporting_post', 'sides_first', 'along_way', 'human_rights_death', 'impossible', 'hit_back', 'slowly']

Topic 9:

['todays_match', 'iran', 'human_rights', 'iranian', 'iranian_football', 'cavallo', 'us_armband', 'news', 'portugal', 'shock_grief']

Topic 17:

['flights', 'airlines', 'fans', 'issued_statement', 'manager_fernando_santos', 'establishment', 'three_countries', 'achraf_hakimi', 'average', 'drink']

Topic 18:

['jakob_jensen', 'happy', 'danilo', 'qatar_also', 'wage', 'played', 'advocates', 'football_governing_body', 'earlier_year', 'return_home']

Topic 19:

['appearing_shadow_domestic', 'took_control', 'agency_tass', 'xherdan_shaqiri', 'vladimir_putin', 'wanted_change_official', 'evening', 'recently_told', 'lifting_trophy', 'us_soccer_f']



FIFA WORLD CUP
Qatar2022

Number of topics $k = 29$ for LDA was chosen with use of coherence.

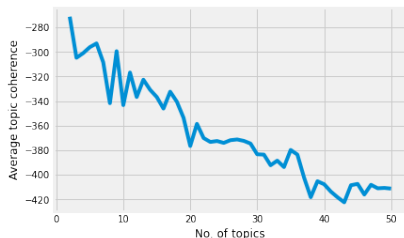


Figure: Coherence for TF

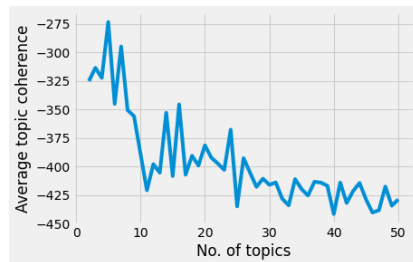


Figure: Coherence for TF-IDF

Parameters Optimization TF Qatar

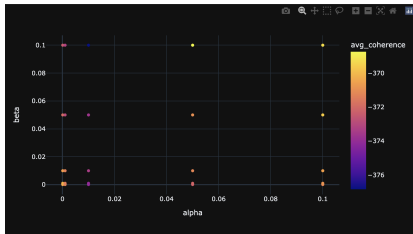


Figure: Plot of avg. coherence vs α and β

| | alpha | beta | vectorizer | avg_coherence |
|----|--------|--------|------------|---------------|
| 38 | 0.0500 | 0.1000 | tf | -368.736847 |
| 48 | 0.1000 | 0.1000 | tf | -369.242746 |
| 46 | 0.1000 | 0.0500 | tf | -369.582713 |
| 2 | 0.0001 | 0.0010 | tf | -370.177741 |
| 0 | 0.0001 | 0.0001 | tf | -370.428636 |
| 4 | 0.0001 | 0.0100 | tf | -370.564547 |
| 44 | 0.1000 | 0.0100 | tf | -370.573238 |
| 42 | 0.1000 | 0.0010 | tf | -370.866007 |
| 36 | 0.0500 | 0.0500 | tf | -370.972093 |
| 40 | 0.1000 | 0.0001 | tf | -371.199647 |
| 34 | 0.0500 | 0.0100 | tf | -371.801170 |

30 rows x 4 columns [Open in new tab](#)

Figure: Table of avg. coherence vs α and β

Conclusion $\alpha = 0.05$, $\beta = 0.1$.

Parameters Optimization TF-IDF Qatar

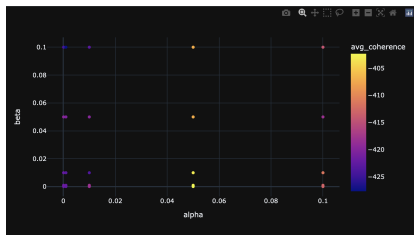


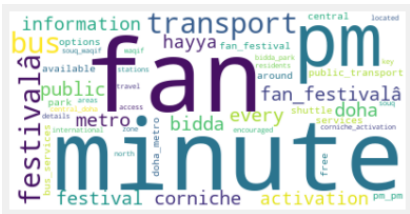
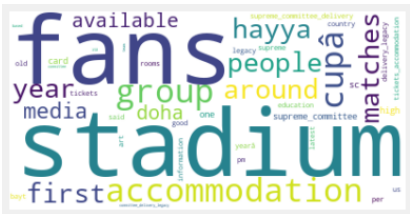
Figure: Plot of avg. coherence vs α and β

| | alpha | beta | vectorizer | avg_coherence |
|----|--------|--------|------------|---------------|
| 33 | 0.0500 | 0.0010 | tf-idf | -402.343844 |
| 31 | 0.0500 | 0.0001 | tf-idf | -402.414570 |
| 35 | 0.0500 | 0.0100 | tf-idf | -402.978331 |
| 37 | 0.0500 | 0.0500 | tf-idf | -407.141852 |
| 39 | 0.0500 | 0.1000 | tf-idf | -407.144554 |
| 45 | 0.1000 | 0.0100 | tf-idf | -411.493241 |
| 41 | 0.1000 | 0.0001 | tf-idf | -413.003530 |
| 43 | 0.1000 | 0.0010 | tf-idf | -413.039402 |
| 49 | 0.1000 | 0.1000 | tf-idf | -414.295343 |
| 23 | 0.0100 | 0.0010 | tf-idf | -418.143389 |

Figure: Table of avg. coherence vs α and β

Conclusion $\alpha = 0.05$, $\beta = 0.001$.

TF Word Clouds Qatar



LDA TF Topics Qatar

```
Topic 0: ['fans', 'stadium', 'cupā', 'first', 'also', 'said', 'during', 'lusail', 'host', 'group']
Topic 1: ['cupā', 'first', 'work', 'like', 'legacy', 'stadium', 'win', 'working', 'middle', 'new']
Topic 2: ['fans', 'part', 'stadium', 'one', 'people', 'volunteer', 'training', 'sc', 'said', 'team']
Topic 3: ['workers', 'rights', 'security', 'continued', 'ensure', 'government', 'contractors', 'human', 'dedicated', 'testament']
Topic 4: ['match', 'set', 'exhibition', 'training', 'looking forward', 'despite', 'team', 'argentina', 'souq', 'several']
Topic 5: ['first', 'stars', 'seen', 'th', 'footballā', 'among', 'edition cupā', 'once in lifetime', 'wonderful', 'ambassadors']
Topic 6: ['stadium', 'fans', 'also', 'accommodation', 'cupā', 'during', 'group', 'people', 'first', 'hayya']
Topic 7: ['every day', 'one', 'thawadi', 'lusail stadium', 'kick off', 'central doha', 'spoke', 'fans', 'legacy sc', 'canā']
Topic 8: ['fans', 'first', 'including', 'doha', 'arab', 'corniche', 'work', 'cupā', 'central', 'support']
Topic 9: ['people', 'supreme committee', 'international', 'accessible', 'argentina', 'cooling', 'seeing', 'plan', 'up round', 'proud']
Topic 10: ['fan', 'during', 'fans', 'around', 'people', 'event', 'legacy', 'music', 'everyone', 'cupā']
Topic 11: ['sustainable', 'yearā', 'thank', 'fans', 'legacy', 'cameroon', 'predictions supreme', 'company', 'gave predictions', 'future']
Topic 12: ['first', 'middle east', 'during yearā', 'trophy', 'team', 're', 'middle', 'east arab', 'country', 'mega event']
Topic 13: ['amazing', 'generation', 'generation amazing', 'social', 'development', 'jordan', 'foundation', 'education', 'youth', 'skills']
Topic 14: ['legacy', 'sc', 'delivery', 'supreme committee', 'supreme', 'supreme committee delivery', 'committee delivery', 'committee delivery legacy', 'delivery legacy', 'committee']
Topic 15: ['audio', 'audio descriptive', 'descriptive', 'commentary', 'youth', 'app', 'descriptive commentary', 'audio descriptive commentary', 'key role', 'blind']
Topic 16: ['said', 'first', 'executive director', 'generation', 'fans', 'international', 'development', 'learn', 'sport', 'skills']
Topic 17: ['blind', 'commentary', 'descriptive', 'descriptive commentary', 'audio descriptive', 'audio', 'audio descriptive commentary', 'arabic', 'sighted fans', 'partially sighted fans']
Topic 18: ['pm', 'fan', 'transport', 'bus', 'festivalā', 'fan festivalā', 'metro', 'doha', 'corniche', 'public']
Topic 19: ['volunteers', 'volunteer', 'morocco', 'positive', 'fan', 'medical', 'said', 'heart', 'predictions supreme', 'cupā']
Topic 20: ['fans', 'us', 'part', 'said', 'very', 'started', 'everyone', 'great', 'weā', 'take part']
Topic 21: ['experience', 'during', 'cultural', 'fans', 'access', 'players', 'four', 'amazing', 'showcase', 'one']
Topic 22: ['hayya', 'international', 'hayya card', 'music', 'around', 'fans', 'transport', 'festival', 'mandatory', 'pm']
Topic 23: ['khori', 'global sustainability', 'predictions supreme committee', 'ronald de', 'years ago', 'initiative', 'plenty', 'rest', 'leave', 'released']
Topic 24: ['ambassadors', 'feels', 'providing', 'generation', 'pride', 'title', 'players', 'fan festival', 'up', 'said']
Topic 25: ['expected', 'metro', 'only days away', 'days away', 'only days', 'bus', 'detailed', 'drive', 'bus services', 'doha metro']
Topic 26: ['legacy', 'people', 'generation', 'amazing', 'sustainability', 'cupā', 'stadium', 'industry', 'quality', 'attended']
Topic 27: ['cupā', 'said', 'many', 'matches', 'stadium', 'iā', 'during', 'middle', 'legacy', 'fans']
Topic 28: ['fan', 'mark', 'centre', 'stadium host fans', 'world', 'fans', 'engagement', 'tournamentā', 'khori executive director', 'volunteering']
```

Topic 3:

['workers', 'rights', 'security', 'continued', 'ensure', 'government',
'contractors', 'human', 'dedicated', 'testament']

Topic 6:

['stadium', 'fans', 'also', 'accommodation', 'cupâ', 'during', 'group',
'people', 'first', 'hayya']

Topic 12:

['first', 'middle east', 'during yearâ', 'trophy', 'team', 're', 'middle', 'east
arab', 'country', 'mega event']

Topic 18:

['pm', 'fan', 'transport', 'bus', 'festivalâ', 'fan festivalâ', 'metro', 'doha',
'corniche', 'public']

TF-IDF Word Clouds Qatar



LDA TF-IDF Topics Qatar

```
Topic 0: ['event', 'available', 'fans', 'pm', 'we', 'also', 'tickets', 'people', 'added', 'opening_match']
Topic 1: ['jordan', 'global_sustainability_assessment', 'recommend', 'human', 'khori_executive', 'thanks', 'generations', 'hamad_international', 'advice_would', 'generation_amazing']
Topic 2: ['hayya', 'process', 'fans', 'benefits', 'help', 'series', 'sc', 'entry_permit', 'act_entry_permit', 'ticket']
Topic 3: ['city', 'safe', 'group', 'senegal', 'wide', 'al', 'lusail', 'boost', 'committee', 'government']
Topic 4: ['artists', 'lusail', 'hilaal', 'fans', 'match', 'sc', 'super', 'rail', 'training', 'enjoyed']
Topic 5: ['volunteer', 'volunteers', 'training', 'learning', 'edition_cupâ', 'public', 'seen', 'face', 'role', 'once_in_lifetime']
Topic 6: ['growing', 'it', 'generate', 'visited', 'around', 'massive', 'everywhere', 'per', 'yearâ', 'successful']
Topic 7: ['airport', 'overall', 'compact', 'hamad_international', 'one', 'previous', 'arriving', 'tournaments', 'water', 'sustainable']
Topic 8: ['describe', 'kind', 'support', 'aims', 'nasser_khori_executive', 'work', 'held_december', 'itâ', 'soon', 'canâ']
Topic 9: ['plan', 'argentina', 'supreme_committee', 'up_round', 'appeared', 'gulf', 'artists', 'people', 'visiting', 'player']
Topic 10: ['looking_forward', 'looking', 'forward', 'originally', 'science', 'appeared', 'footballâ_showpiece_event', 'canada', 'spot', 'currently']
Topic 11: ['sustainable', 'road', 'lusail', 'does', 'yearâ', 'milestone', 'thank', 'taking_place', 'qualified', 'october']
Topic 12: ['during_yearâ', 'ali_stadium', 're', 'welcomed', 'multiple', 'mega_event', 'historic', 'middle_east', 'the_art_stadiums', 'it']
Topic 13: ['development', 'jordan', 'sustainable_development_goals', 'focusing', 'fans_across_globe', 'generation_amazing', 'believes', 'lionel_messi', 'act_entry', 'act']
Topic 14: ['start', 'recently', 'final', 'break', 'easy', 'during', 'interior', 'bus', 'better', 'pleased']
Topic 15: ['watch', 'amazing', 'reached', 'time', 'trip', 'eng_bodour', 'places', 'remember', 'kick', 'formed']
Topic 16: ['hayya', 'card_holders', 'card', 'executive_director', 'feature', 'public', 'non', 'hayya_card', 'follow', 'international']
Topic 17: ['generation_amazing', 'generation', 'amazing', 'jordan', 'youth', 'programmes', 'communities', 'generation_amazing_foundation', 'amazing_foundation', 'schools']
Topic 18: ['stadium', 'janoub', 'janoub_stadium', 'wakrah', 'group', 'stadiumâ', 'stadium_host', 'sustainability', 'venue', 'design']
Topic 19: ['volunteers', 'road', 'training', 'together', 'lusail_stadium', 'lusail', 'staff', 'guests', 'marked', 'learning']
Topic 20: ['native', 'very', 'bodour', 'al', 'great', 'started', 'fans', 'oman', 'four_matches', 'including']
Topic 21: ['blind', 'commentary', 'descriptive', 'descriptive_commentary', 'audio', 'audio_descriptive', 'audio_descriptive_commentary', 'partially_sighted_fans', 'partially_sighted', 'sighted']
Topic 22: ['group', 'morocco', 'around', 'this', 'hayya', 'canada', 'festivalâ_bidda', 'thumama_stadium', 'years', 'poland']
Topic 23: ['khori', 'predictions_supreme_committee', 'global_sustainability', 'ronald_de', 'years_ago', 'initiative', 'leave', 'plenty', 'people', 'rest']
Topic 24: ['providing', 'feels', 'title', 'bayt_stadium', 'fan_festival', 'cultural', 'thousands_fans', 'venue_host', 'teamwork', 'up']
Topic 25: ['exhibition', 'press', 'youth', 'several', 'morocco', 'once', 'human', 'doha', 'seen', 'hosted']
Topic 26: ['expertise', 'awareness', 'programmes', 'generation_amazing', 'share', 'legacy', 'during_yearâ', 'foundation', 'benefit', 'generation']
Topic 27: ['fans', 'stadium', 'hayya', 'lusail', 'during', 'first', 'group', 'cupâ', 'legacy', 'said']
Topic 28: ['ahmad_bin', 'bin', 'ali_stadium', 'bin_ali_stadium', 'bin_ali', 'ahmad_bin_ali', 'lusail', 'ahmad', 'special', 'native']
```

Topic 2:

['hayya', 'process', 'fans', 'benefits', 'help', 'series', 'sc', 'entry
permit', 'act_entry_permit', 'ticket']

Topic 12:

['during_yearâ', 'ali_stadium', 're', 'welcomed', 'multiple', 'mega_event',
'historic', 'middle_east', 'the_art_stadiums', 'it']

Time for BERT



Time for BERT





argentina match group team first qatar france messi goal game



human rights qatar cnn fifa people said workers tournament



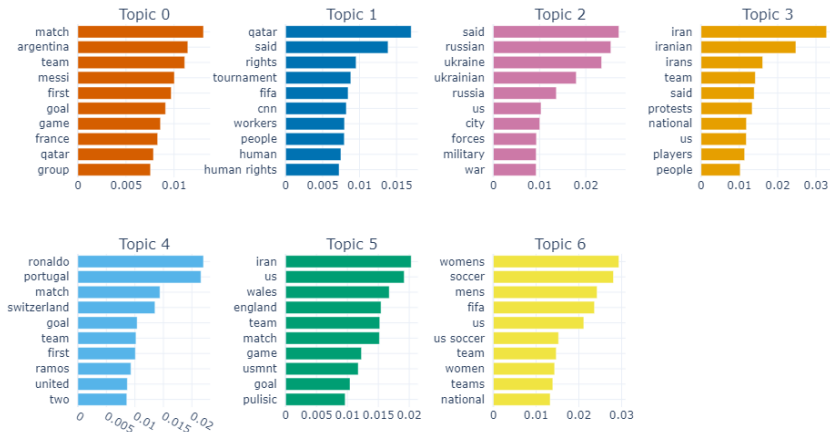
russian said war forces us city military russia ukrainian ukraine



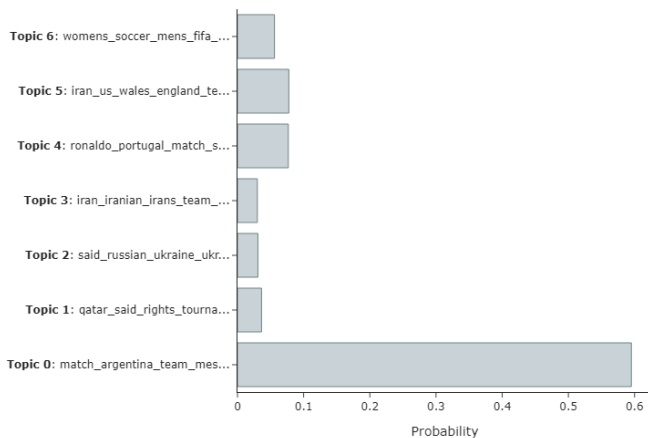
said iran team people protests players national us iranian

BERT CNN

Topic Word Scores



Topic Probability Distribution





A word cloud for BERT Qatar. The most prominent words are 'first', 'east', 'middle', 'fans', 'said', 'region', 'players', 'cup', 'arab', and 'middle'. The words are arranged in a cluster, with 'middle' being the largest and most central.



A word cloud for BERT Qatar. The most prominent words are 'amazing', 'generation', 'work', 'training', 'volunteer', 'people', 'sc', 'legacy', and 'fans'. The words are arranged in a cluster, with 'amazing' and 'generation' being the largest and most central.

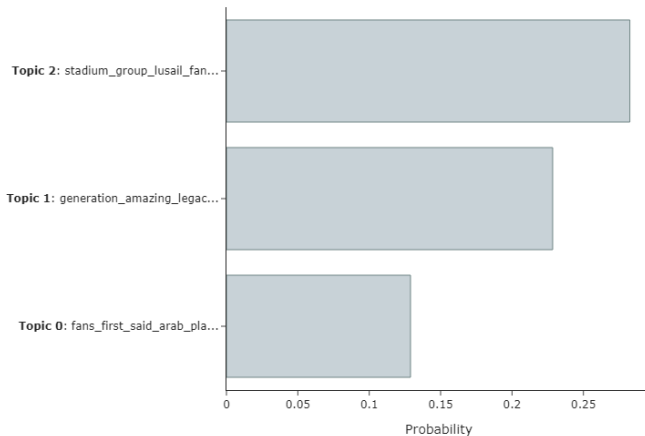


A word cloud for BERT Qatar. The most prominent words are 'lusail', 'stadium', 'fans', 'group', 'city', 'venue', 'matches', 'metro', and 'december'. The words are arranged in a cluster, with 'stadium' and 'group' being the largest and most central.

Topic Word Scores



Topic Probability Distribution



Conclusions

We can clearly see differences in topics presented by West media and Qatar media. Topics such as *human rights* or *Iranian protests* were completely omitted in the Qatari media.

Summary of optimized LDA parameters:

| | CNN | Qatar |
|--------|-------------------------------|--------------------------------|
| TF | $\alpha = 0.1, \beta = 0.01$ | $\alpha = 0.05, \beta = 0.1$ |
| TF-IDF | $\alpha = 0.001, \beta = 0.1$ | $\alpha = 0.05, \beta = 0.001$ |

Further work: analysis of scrapped twitter dataset.



PhD Candidate Maciej Światała

Topic Modelling Masters Course

[https://usosweb.uw.edu.pl/kontroler.php?action =
katalog2/przedmioty/pokazPrzedmiotprz_kod = 2400 – ZEWW878](https://usosweb.uw.edu.pl/kontroler.php?action=katalog2/przedmioty/pokazPrzedmiotprz_kod=2400-ZEWW878)

Thank you