

Name of the topic:

After running LDA on Wikipedia, I got topic names as Topic 0, Topic 1 to Topic 19. Then I have checked the topic structure and found that it is a tuple consist of index number and topic word weight. So I have considered two words with highest weight, concatenated these two words and given as topic name. Else, I have to give topic name manually.

Topic names:

1. party'&'election
2. album'&'song
3. india'&'indian
4. ngc'&'medicine
5. ship'&'aircraft
6. bank'&'business
7. art'&'german
8. television'&'cast
9. physics'&'mathematics
10. canton'&'saint
11. church'&'bishop
12. league'&'football
13. species'&'genus
14. mf'&'game
15. illinois'&'michigan
16. research'&'education
17. greek'&'temple
18. station'&'railway
19. la'&'finals
20. army'&'al

Topics derived:

```
["party'&'election'", '0.021*party' + 0.021*election' + 0.011*assembly' + 0.010*electected' + 0.009*minister' + 0.009*democratic' + 0.009*politician' + 0.009*president' + 0.008*legislative' + 0.007*council']

["album'&'song'", '0.015*album' + 0.010*song' + 0.007*chart' + 0.007*band' + 0.006*track' + 0.005*vocals' + 0.004*guitar' + 0.004*label' + 0.004*singles' + 0.004*listing']

["india'&'indian'", '0.013*india' + 0.012*indian' + 0.008*singh' + 0.008*pradesh' + 0.008*chinese' + 0.008*hong' + 0.007*china' + 0.007*kong' + 0.007*tamil' + 0.006*li']

["army'&'al'", '0.006*army' + 0.005*al' + 0.005*regiment' + 0.004*military' + 0.003*division' + 0.003*pakistan' + 0.003*infantry' + 0.003*mosque' + 0.002*forces' + 0.002*corps']

["ngc'&'medicine'", '0.010*ngc' + 0.008*medicine' + 0.008*protein' + 0.008*medical' + 0.008*cancer' + 0.007*spiders' + 0.006*clinical' + 0.006*gene' + 0.006*disease' + 0.006*treatment']
```

```
["'la'&'finals'", '0.007*la' + 0.006*finals' + 0.006*spanish' + 0.005*reorganisation' + 0.005*el' + 0.005*del' + 0.005*italian' + 0.005*fw' + 0.005*fiba' + 0.004*josé'']

["'ship'&'aircraft'", '0.007*ship' + 0.006*aircraft' + 0.006*racing' + 0.005*air' + 0.005*navy' + 0.005*class' + 0.005*car' + 0.004*ships' + 0.004*engine' + 0.004*motorsport'']

["'bank'&'business'", '0.007*bank' + 0.005*business' + 0.005*products' + 0.005*million' + 0.005*bangladesh' + 0.004*food' + 0.004*companies' + 0.004*founded' + 0.004*investment' + 0.004*market'']

["'station'&'railway'", '0.009*station' + 0.008*railway' + 0.007*building' + 0.007*historic' + 0.006*village' + 0.006*district' + 0.005*river' + 0.005*km' + 0.005*express' + 0.005*road'']

["'art'&'german'", '0.009*art' + 0.006*german' + 0.005*la' + 0.005*le' + 0.005*museum' + 0.005*paris' + 0.005*french' + 0.004*communes' + 0.004*des' + 0.004*der'']
```

```
["'television'&'cast'", '0.003*television' + 0.003*cast' + 0.003*tv' + 0.003*award' + 0.003*directed' + 0.003*festival' + 0.003*episode' + 0.002*films' + 0.002*show' + 0.002*awards'']

["'physics'&'mathematics'", '0.008*physics' + 0.005*mathematics' + 0.004*solar' + 0.004*theory' + 0.004*quantum' + 0.004*galaxy' + 0.004*chemistry' + 0.003*mathematical' + 0.003*attains' + 0.003*research'']

["'canton'&'saint'", '0.057*canton' + 0.045*saint' + 0.026*sun' + 0.020*administrative' + 0.017*effect' + 0.017*seat' + 0.016*france' + 0.012*division' + 0.011*came' + 0.011*french'']

["'church'&'bishop'", '0.011*church' + 0.009*bishop' + 0.005*catholic' + 0.003*london' + 0.003*roman' + 0.003*diocese' + 0.003*william' + 0.003*parish' + 0.003*royal' + 0.003*pope'']

["'league'&'football'", '0.012*league' + 0.010*football' + 0.008*cup' + 0.008*club' + 0.008*tournament' + 0.007*championship' + 0.007*championships' + 0.007*round' + 0.006*player' + 0.006*goals'']
```

```
["'species'&'genus'", '0.022*species' + 0.008*genus' + 0.006*grows' + 0.006*mm' + 0.005*australia' + 0.005*endemic' + 0.005*flowers' + 0.005*description' + 0.005*described' + 0.005*shrub'']

["'mf'&'game'", '0.014*mf' + 0.012*game' + 0.008*software' + 0.006*data' + 0.006*http' + 0.005*www' + 0.005*app' + 0.004*radio' + 0.004*fm' + 0.004*users'']

["'greek'&'temple'", '0.016*greek' + 0.014*temple' + 0.011*ancient' + 0.011*roman' + 0.008*bc' + 0.006*inscription' + 0.006*king' + 0.006*bishopric' + 0.006*ad' + 0.006*cypriot'']

["'illinois'&'michigan'", '0.011*illinois' + 0.009*michigan' + 0.007*texas' + 0.007*virginia' + 0.007*carolina' + 0.007*iowa' + 0.006*baseball' + 0.006*ohio' + 0.006*missouri' + 0.006*washington'']

["'research'&'education'", '0.004*research' + 0.004*education' + 0.003*professor' + 0.003*science' + 0.003*women' + 0.003*institute' + 0.003*law' + 0.002*social' + 0.002*health' + 0.002*president'']
```

Wikipedia links:

1. https://en.wikipedia.org/wiki/US_Lacrosse_Intercollegiate_Associates
2. [https://en.wikipedia.org/wiki/This_Is_My_Country_\(film\)](https://en.wikipedia.org/wiki/This_Is_My_Country_(film))
3. https://en.wikipedia.org/wiki/Ch%C3%A2teau_de_Cordon

Names of the topic assigned to three wiki links:

```

web page1
Score: 0.766636848449707      Topic: 0.004*"research" + 0.004*"education" + 0.003*"professor" + 0.003*"science" + 0.003*"women" + 0.003*"institute" + 0.003*"law" + 0.002*"social" + 0.002*"health" + 0.002*"president"
Score: 0.1350073218345642      Topic: 0.007*"ship" + 0.006*"aircraft" + 0.006*"racing" + 0.005*"air" + 0.005*"navy" + 0.005*"class" + 0.005*"car" + 0.004*"ships" + 0.004*"engine" + 0.004*"motorsport"
Score: 0.07722198218107224      Topic: 0.006*"army" + 0.005*"al" + 0.005*"regiment" + 0.004*"military" + 0.003*"division" + 0.003*"pakistan" + 0.003*"infantry" + 0.003*"mosque" + 0.002*"forces" + 0.002*"corps"
Score: 0.011850272305309772      Topic: 0.011*"church" + 0.009*"bishop" + 0.005*"catholic" + 0.003*"london" + 0.003*"roman" + 0.003*"diocese" + 0.003*"william" + 0.003*"parish" + 0.003*"royal" + 0.003*"pope"

web page2
Score: 0.710037112236023      Topic: 0.004*"research" + 0.004*"education" + 0.003*"professor" + 0.003*"science" + 0.003*"women" + 0.003*"institute" + 0.003*"law" + 0.002*"social" + 0.002*"health" + 0.002*"president"
Score: 0.14069128036499023      Topic: 0.006*"army" + 0.005*"al" + 0.005*"regiment" + 0.004*"military" + 0.003*"division" + 0.003*"pakistan" + 0.003*"infantry" + 0.003*"mosque" + 0.002*"forces" + 0.002*"corps"
Score: 0.04786279797554016      Topic: 0.003*"television" + 0.003*"cast" + 0.003*"tv" + 0.003*"award" + 0.003*"directed" + 0.003*"festival" + 0.003*"episode" + 0.002*"films" + 0.002*"show" + 0.002*"awards"
Score: 0.02692112699151039      Topic: 0.010*"ngc" + 0.008*"medicine" + 0.008*"protein" + 0.008*"medical" + 0.008*"cancer" + 0.007*"spiders" + 0.006*"clinical" + 0.006*"gene" + 0.006*"disease" + 0.006*"treatment"

2019-03-23 12:22:14,431 : INFO : adding document #0 to Dictionary(0 unique tokens: [])
2019-03-23 12:22:14,434 : INFO : built Dictionary(157 unique tokens: ['p', 'b', 'château', 'de', 'cordon']...) from 341 documents (total 341 corpus positions)

Score: 0.022435219958424568      Topic: 0.021*"party" + 0.021*"election" + 0.011*"assembly" + 0.010*"elected" + 0.009*"minister" + 0.009*"democratic" + 0.009*"politician" + 0.009*"president" + 0.008*"legislative" + 0.007*"council"
Score: 0.016937196254730225      Topic: 0.007*"la" + 0.006*"finals" + 0.006*"spanish" + 0.005*"reorganisation" + 0.005*"el" + 0.005*"del" + 0.005*"italian" + 0.005*"fw" + 0.005*"fiba" + 0.004*"josé"
Score: 0.013441824354231358      Topic: 0.007*"ship" + 0.006*"aircraft" + 0.006*"racing" + 0.005*"air" + 0.005*"navy" + 0.005*"class" + 0.005*"car" + 0.004*"ships" + 0.004*"engine" + 0.004*"motorsport"

Score: 0.013441824354231358      Topic: 0.007*"ship" + 0.006*"aircraft" + 0.006*"racing" + 0.005*"air" + 0.005*"navy" + 0.005*"class" + 0.005*"car" + 0.004*"ships" + 0.004*"engine" + 0.004*"motorsport"
Score: 0.013192732818424702      Topic: 0.016*"greek" + 0.014*"temple" + 0.011*"ancient" + 0.011*"roman" + 0.008*"bc" + 0.006*"inscription" + 0.006*"king" + 0.006*"bishops" + 0.006*"ad" + 0.006*"cypriot"

web page3
Score: 0.7040810585021973      Topic: 0.004*"research" + 0.004*"education" + 0.003*"professor" + 0.003*"science" + 0.003*"women" + 0.003*"institute" + 0.003*"law" + 0.002*"social" + 0.002*"health" + 0.002*"president"
Score: 0.129910409450531      Topic: 0.003*"television" + 0.003*"cast" + 0.003*"tv" + 0.003*"award" + 0.003*"directed" + 0.003*"festival" + 0.003*"episode" + 0.002*"films" + 0.002*"show" + 0.002*"awards"
Score: 0.11481393128633499      Topic: 0.006*"army" + 0.005*"al" + 0.005*"regiment" + 0.004*"military" + 0.003*"division" + 0.003*"pakistan" + 0.003*"infantry" + 0.003*"mosque" + 0.002*"forces" + 0.002*"corps"
Score: 0.022828221321105957      Topic: 0.010*"ngc" + 0.008*"medicine" + 0.008*"protein" + 0.008*"medical" + 0.008*"cancer" + 0.007*"spiders" + 0.006*"clinical" + 0.006*"gene" + 0.006*"disease" + 0.006*"treatment"
Score: 0.013361331075429916      Topic: 0.007*"la" + 0.006*"finals" + 0.006*"spanish" + 0.005*"reorganisation" + 0.005*"el" + 0.005*"del" + 0.005*"italian" + 0.005*"fw" + 0.005*"fiba" + 0.004*"josé"

```

Observation:

First web page is related to a formation of team for the sports Lacrosse and topic 16(research & education) got highest score. Second link is about a Filipino drama film. Second link also got topic 16 as a highest scoring topic. Third link is about a historical region . It also got topic 16 as high scoring topic. I observed that one topic has more inclination towards three links. But I don't why like that happened.