**상황**

- BERTopic - 가장 기본 모델
- 지금까지 대강 분류한 10500여개의 특허 데이터
- 관사, 조사, 전치사, 불용어 등 제거 없이 모든 텍스트
- 단순히 BERTopic만 돌렸을때

---

```
     Topic  Count                                             Name  \
0       -1   2993                                 -1_of_method_to_for
1        0    206                   0_aerial_unmanned_aircraft_flight
2        1    175                1_neural_deep_learning_convolutional
3        2    145             2_traffic_flow_information_surveillance
4        3    141          3_logistics_logistical_management_reverse
..     ...    ...                                               ...
236    235     11                           235_tcp_offload_host_can
237    236     11            236_trusted_isp_reliable_trustworthiness
238    237     11          237_coordinate_delivery_timing_optimizing
239    238     11                 238_module_vehicle_node_automobile
240    239     10  239_networksensitive_collection_industrial_alt...
```

위의 내용이 결과 정보인데, 각 행이 하나의 토픽이고
Topic '-1'은 "잡음"을 뜻함. => 관사, 조사, 불용어들로 문서가 너무 많이 날라갔음.
미리 전처리 과정을 통해서 없애고 하면 더 잘될듯

---

**'명칭(제목)'열으로 돌렸을 때**

```
제목열로 돌리니깐 주제가 쓸데없는 말이 적어서 주제가 명확하길래 돌려봤어여
```

### 1번 토픽 - 205

[('aerial', 0.06845507769724539), ('unmanned', 0.06393348937976026), ('aircraft', 0.054788621048037736), ('flight', 0.033411148957681464), ('drone', 0.031467021419405966), ('electric', 0.023219696199290613), ('vehicles', 0.01885707555621402), ('vehicle', 0.015592271711093031), ('uav', 0.015139069646387504), ('landing', 0.015062154603369597)]

### 2번 토픽 - 175

[('neural', 0.0957768986677105), ('deep', 0.04775748662850265), ('learning', 0.04041644164926208), ('convolutional', 0.03648856088189897), ('networks', 0.03162573887245685), ('compute', 0.031015276749614122), ('mechanism', 0.03054128030902052), ('training', 0.02302745831716863), ('machine', 0.017308895749014837), ('optimization', 0.017237394045243822)]

### 3번 토픽 - 145

[('traffic', 0.09422720360882055), ('flow', 0.015562699861874587), ('information', 0.01547042050548071), ('surveillance', 0.014655177611064121), ('analyzing', 0.014252056885462312), ('regions', 0.013409990135476195), ('program', 0.01203056712539087), ('providing', 0.011683469333324458), ('adsb', 0.011213421639604624), ('geographic', 0.01105246215088178)]

### 4번 토픽 - 141

[('logistics', 0.10609087600037088), ('logistical', 0.030132637730290204), ('management', 0.016467117629164602), ('reverse', 0.015511808135311089), ('support', 0.0125418585656959), ('managing', 0.012412931416849087), ('transmodal', 0.012351696822819521), ('logistic', 0.011890509632358047), ('certificate', 0.011633856101483316), ('production', 0.010750164484882198)]

### 5번 토픽 ...

[('security', 0.033878111683326585), ('malicious', 0.026310827001807008), ('attack', 0.02398622078947292), ('against', 0.023387401779384004), ('malware', 0.020412356261795713), ('attacks', 0.019678873540578762), ('protection', 0.017209503177517568), ('threats', 0.017131532008586775), ('detecting', 0.015412021201974116), ('compromise', 0.014899335453094938)]

### 6번 토픽 ...

[('maritime', 0.061686286416078685), ('vessel', 0.06025492336273685), ('vessels', 0.03644310546207928), ('marine', 0.03240673047588115), ('ais', 0.017338595060423984), ('extracting', 0.016905734860330587), ('supervisory',

0.016409428749189416), ('ship', 0.015441477459572465), ('navigation', 0.015344100258535092), ('monitoring', 0.014417497490199967)]

---

**'요약열'으로 돌렸을때**

### 1번 토픽

[('structured', 0.03278573195161213), ('circuit', 0.03231728008105728), ('industrial', 0.025863680283598083), ('values', 0.02573794748959598), ('collection', 0.023857047037523317), ('detection', 0.023694918500249885), ('collector', 0.023154377831049903), ('channels', 0.020566804973442928), ('plurality', 0.02024659056376829), ('input', 0.017610844864015318)]

### 2번 토픽

[('vehicle', 0.032295746218856294), ('module', 0.01185142815207883), ('bus', 0.01083662695943735), ('vim', 0.009100077883964906), ('vehicular', 0.008520525406015002), ('communications', 0.008291506338270015), ('diagnostic', 0.00791862791053952), ('communication', 0.007444290948730353), ('invehicle', 0.007372718200225609), ('vehicles', 0.0070637585211812135)]

### 3번 토픽

[('pressure', 0.04340503176146014), ('fluid', 0.043374652465340294), ('valve', 0.025040365337480126), ('liquid', 0.020559976113351343), ('vessel', 0.01660892997565576), ('gas', 0.015761559117236246), ('conduit', 0.014883342967078722), ('flow', 0.01446060866135438), ('tank', 0.01110940205264618), ('water', 0.010768996016777574)]

### 4번 토픽

[('blood', 0.040039195010992124), ('image', 0.02758679764680411), ('vessel', 0.016938342028630222), ('images', 0.01653389149683792), ('ultrasound', 0.0159215981580468), ('anatomical', 0.015409946065653319), ('geometric', 0.01409029190913449), ('vascular', 0.012982900739365035), ('imaging', 0.012946121547801356), ('tissue', 0.012013689708308609)]

### 5번 토픽

[('security', 0.02825949427872398), ('container', 0.026971019365031955), ('locking', 0.017408887044673273), ('closure', 0.01585458336245687), ('cargo', 0.014827009172160129), ('chassis', 0.012386238865934773), ('door', 0.011537351774568232), ('tamper', 0.009337557539199683), ('threat', 0.009324009778477667), ('lock', 0.008816904728233222)]

---