

Emerging Technologies in the Last Decade

BAN432 Applied Textual Data Analysis for Business and Finance

Final Project – Fall 2019

Lecturers: Christian Langerfeld, Maximillian Rohrer

Candidate no: 6, 10, 55, 90, 91



An Overview of Applied Methodologies



SEC's industry classification

- Gather industry affiliation of the IPO companies
- Gives a first indication of emerging industries, technologies and its keywords



Keyness and KWIC Analysis

- Identify relevant keywords by comparing the study corpus with a reference corpus
- Determine the relevance of the keywords with a keyword in context analysis



Topic model

- Based on a word list from the keyness analysis estimate a topic model
- Identify topics for further insights into technological trends

SEC's Industry Classification

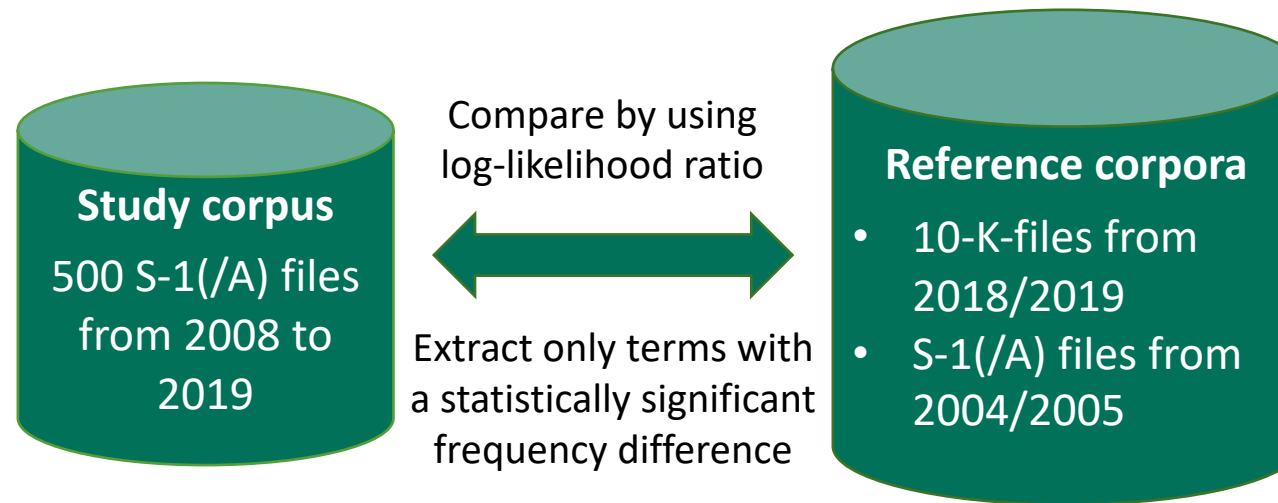
A first indication of emerging industries

| | Industry Description | Frequency |
|-------------------|--|-----------|
| | Pharmaceutical Preparations | 157 |
| | Services-Prepackaged Software | 80 |
| | Blank Checks | 67 |
| Services-Business | Services, Not Elsewhere Classified | 55 |
| | Crude Petroleum and Natural Gas | 50 |
| | Metal Mining | 44 |
| | State Commercial Banks | 40 |
| | Services-Computer Processing and Data Preparation | 35 |
| | Biological Products, (No Diagnostic Substances) | 31 |
| | Surgical and Medical Instruments and Apparatus | 29 |
| | Services-Computer Programming, Data Processing, Etc. | 22 |
| | Services-Management Consulting Services | 22 |
| | Savings Institution, Federally Chartered | 21 |
| | Retail-Miscellaneous Retail | 17 |
| | Gold And Silver Ores | 16 |

Keyness and KWIC Analysis

Step 1: Identify relevant keywords

- Comparing textual information of new companies with old companies will inform us about new emerging technologies



- After setting some bounds, we obtained a list of about 150 keywords
- Based on this list, we have selected 20 keywords that stand out

Keyness and KWIC Analysis

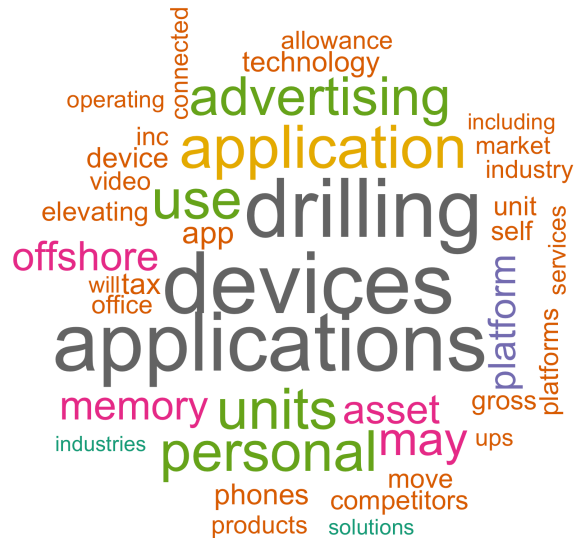
Select 10 relevant keywords for the further KWIC analysis

| | Trend Word | Freq Reference Corpus | Freq Study Corpus | LL |
|----|---------------|-----------------------|-------------------|------------|
| 1 | hospital | 169 | 34386 | 36972.5943 |
| 2 | drilling | 352 | 14080 | 13179.3863 |
| 3 | gaming | 722 | 8522 | 5768.0910 |
| 4 | cloud | 2937 | 286 | 3330.4048 |
| 5 | oil | 2993 | 9289 | 1880.4759 |
| 6 | physician | 188 | 2423 | 1700.4484 |
| 7 | clinical | 2026 | 6558 | 1429.7894 |
| 8 | cars | 145 | 1897 | 1339.1682 |
| 9 | digital | 2367 | 814 | 1281.1319 |
| 10 | cyber | 1400 | 301 | 1106.1743 |
| 11 | semiconductor | 1070 | 166 | 1011.5250 |
| 12 | cybersecurity | 938 | 106 | 1011.2111 |
| 13 | networks | 1847 | 648 | 979.3932 |
| 14 | devices | 2312 | 1011 | 945.8447 |
| 15 | analytics | 823 | 104 | 850.4676 |
| 16 | household | 120 | 1300 | 846.9850 |
| 17 | platforms | 1317 | 434 | 743.6182 |
| 18 | automotive | 1033 | 341 | 582.2867 |
| 19 | statistics | 182 | 1171 | 559.6243 |
| 20 | mobile | 2026 | 1256 | 457.4796 |

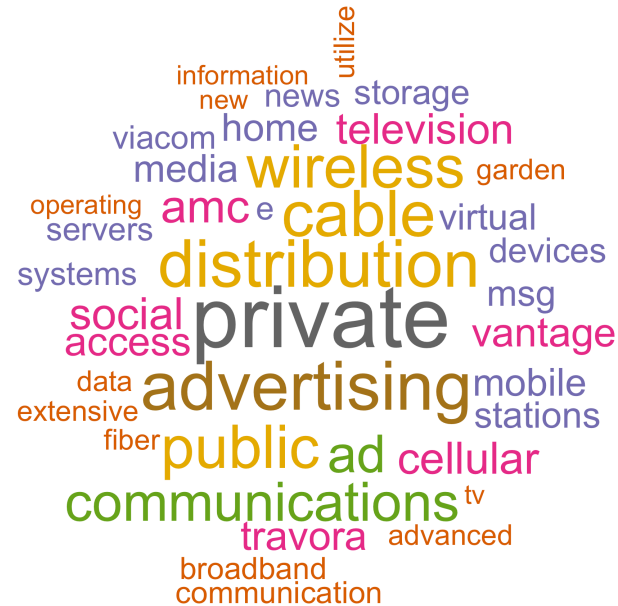
Keyness and KWIC Analysis

Step 2: Look at word clouds to determine the relevance of the keywords

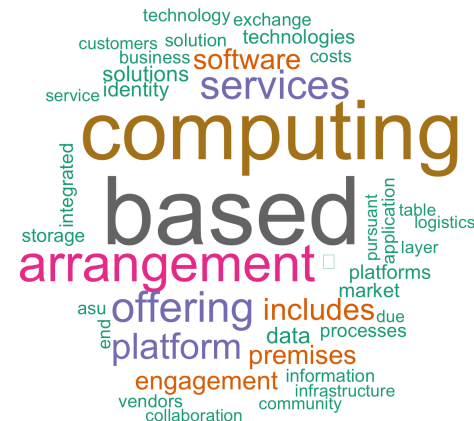
mobile



networks



cloud



automotive



Keyness and KWIC Analysis

Step 3: Interpret top emerging technologies



Mobile & Digital transformation

→ Shift towards mobile applications like apps and digital transformation



Networks & Platforms

→ New cellular technologies and digital platforms, e.g. 5G, IoT



Cloud & Analytics

→ Cloud services and demand for analytics in times of big data



Automotive, Oil & Semiconductor

→ Shift towards sustainability and digitalization influences traditional industries

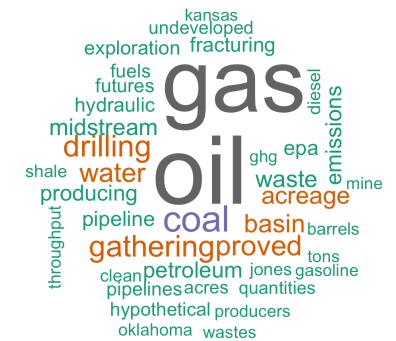
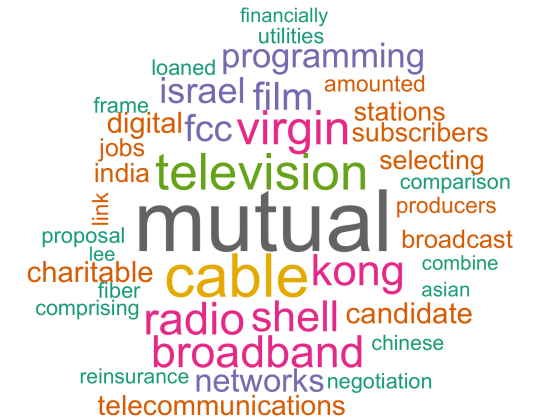


Healthcare innovations

→ Increased use of personal medical devices and raising importance of healthcare in the aging society

Topic Model

- Used only those words which have a statistically significant frequency difference compared to the reference corpus (based on the keyness analysis)
- In total 12 topics estimated
- Four unambiguous topics (the remaining topics are related to finance, gaming, manufacturing, households, etc.)
 - Digitalization
 - Telecommunication
 - Healthcare
 - Oil
- Note: The topic model might not be the most convincing approach, rather a supplement to the previous analysis

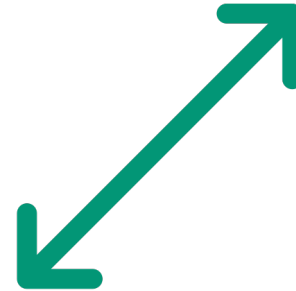


Summary and Limitations



Summary

- We found 10 emerging technologies in the last decade among the IPO companies through the analysis of keyness and KWIC
- Topic model estimated for a further identification and clustering of topics for the IPO companies



Limitations

- Textual data analysis cannot identify important technologies exactly, rather it gives a broad overview about emerging industries and technology fields
- Topic model might not be the best approach because it is unclear how well it works