AI Private Equity Investment: Entrepreneurship Profession Assessment Research

Project Pharaoh
Master of Science in Financial Mathematics Program
Hong Kong University of Science and Technology
Clear Water Bay, Kowloon
aifin.hkust@gmail.com

Abstract

In this project, we did a research on the entrepreneurs in WorkFusion to make an assessment on WorkFusion and the managers based on their expertise and management outcome. We analyze the public speech voice utilizing Google AI and Machine learning technology Voice-to-Text application to convert the speech into text and analyze the text we got with the help of Python programming.

1 Reflection on an AI article--Text-Adaptive Generative Adversarial Networks: Manipulating Images with Natural Language

After reading this paper, I learned a new method of manipulating images using natural language description which is the text-adaptive generative adversarial network (TAGAN) to generate semantically manipulated images while preserving text-irrelevant contents.

In this paper, firstly, the Author proposed that the demand for manipulating or editing images is growing, to make photos to look better or to meet a user's need. Although many techniques have been developed for image editing, editing images specifically per user's intention is still difficult. To solve the problem of existing methods produce reasonable results, but failing to preserve text-irrelevant contents such as the background of the original image. The author introduces their method accurately manipulates images according to the text while preserving text-irrelevant contents.

Secondly, some basic concepts are explained such as GAN structure, Text-Adaptive Generative Adversarial Network (TAGAN), Text-adaptive discriminator, Generator and so on, Simply The method principle is also introduced, and related of the algorithm are given. So I learned some new words about image editing.

Finally, the author did experiments and evaluate by comparing their method to two baseline methods:SISGAN and AttnGAN. they collected a total of 4,000 samples from 20 workers for the user study and set two criterions referred as Accuracy and Naturalness. Average ranking values shows that the new method significantly outperforms baseline methods.

As for the knowledge I learned, I think it give me a new way to come into contact with the image editing area of artificial intelligence. The novel research can really enrich people's daily life. For example, if this method becomes popular, we can easily edit the pictures specially on a customized purpose or a part we intend to change.

2 Reflection on the paper Multifactor Explanations of Asset Pricing Anomalies

This paper mainly introduces anomalies that are not explained by CAPM model. This paper also argues that many of the CAPM average-return anomalies are related, and they are captured by the three-factor model in Fama and French as follows.

$$R_i - R_f = \alpha_i + b_i\left(R_M - R_f\right) + s_i SMB + h_i HML + \varepsilon_i$$

SMB represents the difference between the return on a portfolio of small stocks and the return on a portfolio of large stocks, and HML represents the difference between the return on a portfolio of high book-to -market stocks and the return on a portfolio of low book-to market stocks.

The three-factor model seems to capture much of the cross-sectional variation in average stock returns. Specifically, it captures the return to portfolios formed on E/P, C/P, and sales growth. In general, low E/P, low C/P, and high sales growth are typical of strong firms that have negative slopes on HML. Since HML return is strongly positive, it implies lower expected return and vice versa. And also, stocks with low long-term past returns tend to have positive SMB and SMB return is also positive, and long-term past winners tend to have negative SMB. In total, FF model reveals the reversal of long-term return. However, the continuation of short-term returns is left unexplained by the model. The empirical analysis shows that When the preceding year is included, short-term continuation offsets long-term reversal, and past loser have lower future returns than past winners for portfolios formed with up to four years of past returns.

This paper also put the argument that the realized excess return on any three multifactor-minimum-variance (MMV) portfolios perfectly describe the excess returns on other MMV portfolios. And then prove that different triplets of market portfolio M, small-stock portfolio S, low-B/E portfolio L, and high B/E portfolio H provide equivalent descriptions of returns as FF model.

However, there are still some skepticism about the premium for distress (average HML return). These skepticisms include survivor bias, data snooping, investor over-reaction, and etc.

3 Analyze the Capabilities of Senior Managers

LinkedIn provides information about senior managers in a company, such as academic qualifications, work experience and so on. Here, we mainly focus on the skills recognition in LinkedIn. Skills recognition is the recognition of a number of peers by a manager's ability in a particular field. It can be represented by a numerical value. For example, Alex Lyashok's Software Development is recognized by 127 peer experts, and the Software Development skill can be recorded as 127 skill points.

An executive has many skills. In order to easily show data, we can combine similar skills into one skill. For example, combining business strategy, business process and business transformation into business ability. Then classify the combined skills into technical skills, management skills and leadership skills. The plots of them are shown in Figure 1:

We can find that technical skills are Mainly concentrated on software, management skills are Mainly concentrated on marketing and leadership skills are Mainly concentrated on leadership, partnership and relationship management.

Then, Figure 2 shows ratios of these three type skills to the total skills. As can be seen from the distribution on the plots, the executives of WorkFusion have more recognitions at the technical level, such as Software, accounting for more than half of the total skills points, indicating that the senior management team has certain advantages in terms of technology. On the contrary, leadership is less and leadership is not widely recognized. There are some differences in technology, management and leadership. In order to better reflect this result, we chose WorkFusion's competitor UiPath executives' skills recognition as a comparison. The total skill recognition points are similar. Compared to WorkFusion, UiPath executives' skill recognition distribution is more balanced, and technology, management and leadership are all have a certain degree of recognition.

In addition, we can find that each senior manager has a different emphasis on skills. In Figure 3, rings from inside to outside refer to technical skills, management skills and leadership. Some have more recognition in terms of technical capabilities, and some have more recognition in management. For example, the technical skills of WorkFusion's CEO accounts for 22.4209% of the company's technical competency skills points, which is more than other executives, but his management and leadership ratio is very small. It is the same as CFO Peter Cumello ,but it was small in terms of technology. Of course, there are also managers with high skill certifications, such as Todd Rathje, whose ratios are 13.6685%, 27.3465% and 53.4483% in terms of technology, management and leadership. There are 4 people who have a certain proportion of skills certification in three aspects, accounting for 34.36% . This number is small compared to the competitor UiPath, which has a ratio of 57.1429%. This means that WorkFusion is not so balanced in terms of skill authentication compared to UiPath.

4 Executive Team's Background

Through the WORKFUSION official website, I collected executive team's background information and company major news release information. The background information of senior executives includes three aspects: academic qualifications (bachelor, master, and phd) as well as major job matching situation and the distribution of industry where the former companies belong.

WORKFUSION has 14 executives in total, they are CEO & President Alex Lyashok, CSO Max Yankelevich, CFO Pete Cumello, CMO Sam Fahmy, COO Monica Jonas, CPO Teresa Thomas, CRO Sudip Mitra, Secretary Ginger Mosier, Customer CTO Andrew Volkov, Engineering CTO Marc Ache, Product SVP Mikhail Abramchyk, European Operations SVP Ilya Kazimirovskiy, Professional Services SVP Alexey Vitashkevich, and VP Abby Levenberg.

I have counted the academic qualifications of senior executives. Among them, 8 keep undergraduate degrees, 5 keep master's degrees, and 1 keeps a doctoral degree. In terms of professional counterparts, there are 80% of the executives whose majors during the school match the current jobs. For example, CEO Alex Lyashok obtained MBA degree in Stern School of Business in New York University, and CPO Teresa Thomas studied at Portland. State University, majoring in Human Resources and Business Administration. However, a small number of executives are not professional, for example, COO Monica Jonas studied Comparative Literature during her undergraduate degree.

The industries where executives used to serve contain Software, Media, Management Consulting, Marketing Communication, Government agency, Data Analysis, Clothing and etc. The details can be seen in table 1.

On average, each executive worked in three or more different industries before WORKFUSION. Among them, CSO Max Yankelevich has the most extensive experience. He has been in business for 6 years and has served in 6 different industries, and mainly served in Information technology services company. Sudip Mitra is a relatively younger and less experienced executive, who previously served only IBM and worked for 7 years. Besides, executives work for over 20 years on average.

5 Significant news and product release information

I divide this type of information into three categories: Product Release, Awards/Recognized, Strategic, cooperation/milestone/financing. The timelines of these news can be seen in figure 4, 5, 6.

There are some conclusions of the information：
1.The company will have a product update or a new product release every two months to half a year.
2. In the approximately three years from 2016 to 2019, the company and senior executives received 12 awards and nominations.

3. The company has maintained cooperation with well-known companies in the expansion stage

to build an excellent business ecosystem.

6 Measure the leadership of Workfusion from papers, articles etc.

To analyze the leadership of Workfusion, this part is to measure the creativity of employees which can be reflected by the papers, news, articles, projects and events.

6.1. Number of news articles, events and investors among competitors

To better show the situation on the company of Workfusion. We compare the results with Workfusion's competitors: Thoughtonomy, UiPath and Kryon Systems.

As we can see from Figure 7 in appendix, Workfusion has the second largest number of News, Articles and Events just below that of UiPath. At the same time, the number of News and Articles are nearly 8 times than that of Thoughtonomy.

In addition, another indicator called Crunchbase Rank can reflect the four companies' whole impact as well.

Crunchbase Rank is determined by an algorithm that takes into account the number of connections of a profile within the platform, the amount of community engagement, funding events, news articles, acquisitions, and more. The benefit of Crunchbase Rank is it lets users prioritize their search results by influence. The higher the Rank, the more influential the profile is compared to its peers. The lower the Rank, the less important the profile is to the community. CB rank of Workfusion, Thoughtonomy, UiPath and Kryon Systems are 1004, 56359, 671 and 1904(can be seen in appendix charts: Table 2)

6.2. Number of news/articles/patents/projects/papers among Workfusion.

We can see that from Figure 8 their Vice President Abby Levenberg in data science field has the most numbers of papers. Max Yankelevich, Co-founder & Chief Strategy Officer of workfusion has the most numbers of News/Articles. Ilya Kazimirovskiy, their Senior Vice President in European Operations has the most numbers of Patents/projects, so the members of current leader team in workfusion have the creativity in some extent because they have diverse types such as news, articles, papers, patents and projects.

6.3. Patents and papers citations

From the Figure 9 and 10 in appendix, no matter patents or papers citations, both show that Workfusion indeed has excellent talents in academic and technology fields, but the number of talents (3 people) is very few.

From the Figure 11 in appendix, for CB rank for the team members in Workfusion, Approximate half of team members' ranks are within top 10000. Therefore, the current leadership of Workfusion has influential impact on society in some extent.

7 Entrepreneur Speech Video Recognition, Textuality and Analysis

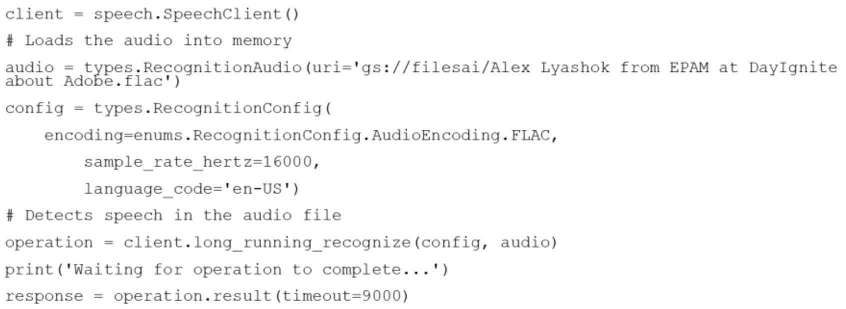
We hope to create a new set methodology to evaluate the profession and innovation of the leaders especially these CEOs etc. that have decision priority and can steer the future path the development of our target corporate. We found the current and formal 1eaders have several good interview and speech videos available on YouTube and some other social medias. For example, Alex Layshok made some speeches on behalf of WorkFusion at forums or some other events like NILF.

There are several widely used machine learning platforms who have got APIs in voice to text, for example IBM, Google and Baidu. After comparing the availability and difficulties, we finally chose Google Cloud Platform as the API library provider.

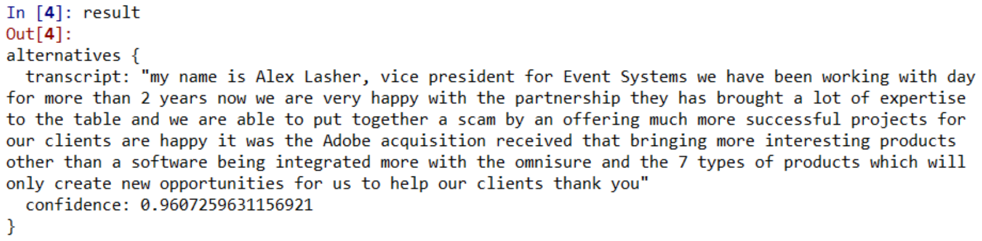
We prepared 11 speech and interview videos. By utilizing the online video converter, we convert the online videos into flac files. Then we are going to take the flac files into our voice to text and opinion analysis program. Before I import the files to the Google Cloud API, I need to check the frequency of flac file is 16000Hz and the type of sound is monotone. Google only supports monotone type of files.

The Google Speech-to Text conversion is powered by machine learning and available for short-form or long-form audio. Google Cloud Speech-to-Text enables developers to convert audio to text by applying powerful neural network models in an easy-to-use API. The API recognizes 120 languages and variants to support your global user base. You can enable voice command-and-control, transcribe audio from call centers, and more. It can process real-time streaming or prerecorded audio, using Google’s machine learning technology. So, it is a great for our purpose to get a text version of the speakers’ speech.

Here is the core part of get our flac file into text strings. The full version will be available in the appendix.



```
client = speech.SpeechClient()
# Loads the audio into memory
audio = types.RecognitionAudio(uri='gs://filesai/Alex Lyashok from EPAM at DayIgnite
about Adobe.flac')
config = types.RecognitionConfig(
    encoding=enums.RecognitionConfig.AudioEncoding.FLAC,
        sample_rate_hertz=16000,
        language_code='en-US')
# Detects speech in the audio file
operation = client.long_running_recognize(config, audio)
print('Waiting for operation to complete...')
response = operation.result(timeout=9000)
```



```
In [4]: result
Out[4]:
alternatives {
  transcript: "my name is Alex Lasher, vice president for Event Systems we have been working with day
for more than 2 years now we are very happy with the partnership they has brought a lot of expertise
to the table and we are able to put together a scam by an offering much more successful projects for
our clients are happy it was the Adobe acquisition received that bringing more interesting products
other than a software being integrated more with the omnisure and the 7 types of products which will
only create new opportunities for us to help our clients thank you"
  confidence: 0.9607259631156921
}
```

And the Voice-To-Text result come in two parts, translation and confidence level. In the translation part, we have the text that we got from the speech, and in the confidence part, it is the confidence level of each sentence we got, which means in what probability our every translation is correct.

After we got the text of the speech, you can make some revise if you want to make no error happens. In our circumstances, we use the text directly this time for a demonstration. We make an analysis framework about the speech of the leader based on the frequency of the words and the selection of their words in different circumstances to evaluate the abilities of speech, ideas they are mainly taking and the theories they believe, so that we will got an good view of their entrepreneurship and professional skills.

First, we divide the transferred texts into words by "nltk" and do preprocessing on them, removing those symbols and worthless words, such like "and", "I", and some prepositions.

To have a direct and abstract view of the speeches, we make wordclouds charts for every text. Take Figure 22 for example, we can see that those words relative to AI & Fintech topics, such as machine learning, artificial intelligence, structuring, crowd-sourced, and automation are more frequently mentioned by the speaker, which is consistent with our expectation.

To understand more precisely how each text is related to the topic and then assess their professionality, we use TF-IDF algorithm to extract key information from the texts. While the wordclouds only consider the frequency of a word in a single text, TF-IDF algorithm takes the whole samples into account, namely, a word appears in more sample texts should be assigned smaller weight since it is not unique enough to reflect the quality of a speech. We also use Sentiment Analysis to assess the positivity and objectivity of these speeches and speakers. We choose "sentiwordnet" in "nltk" as our lexicon, since it already assigns scores to each word from the perspective of positivity, negativity and objectivity respectively. After scoring all samples, we divide the score by length of each text to eliminate the effect of dimension.

Figure 23 shows the scores and ranks of 11 sample texts from perspectives above. When assessing the professionality of texts by TF-IDF algorithm, we select a group of key words relative to AI & Fintech topic to score our samples. And since those words are minor part of the whole texts, the longer the text is, the lower the score of that text. We then set a group of adjusted weights according to the length of texts: $Adjusted\ weight = 1 + floor\left(\frac{word\ number\ of\ text}{500}\right)$, ($floor(x)$ denotes the max integer smaller than x). Note that you can change the key words group and adjusted weights, and even assign different weights to different key words in this method. And the performance of this algorithm will be definitely improved when more samples are collected.

After getting all scores, we normalize them to eliminate dimension among different perspectives by $normalized\ score = \frac{score\ of\ each\ sample}{average\ score\ of\ all\ samples}$. Let Pro, Pos, Obj denote the normalized professionality, positivity (positivity − negativity), and objectivity score respectively. We calculate the weighted total score by $Total\ score = \lambda_1 \times Pro + \lambda_2 \times Pos + \lambda_3 \times Obj$. We set $\lambda_1 = 0.4, \lambda_2 = 0.3, \lambda_3 = 0.3$ here since we believe that the professionality of senior executives of an AI company is more important. You can decide your own weights based on other arguments.

Based on our model and assumptions, we conclude that Max shows the highest professionality, Alex is the most positive when talking and the speeches given by Adam are most objective. In total, Max is the best executive based on his speeches.

And by collecting more samples, we can get more reasonable results. This method can also be used when the number of companies and persons we research on increases, since most of the work can be done by computers

8 Synthesis and Suggestions

The research provides an evaluation of WorkFusion's managers based on their personal technical skills and capabilities, professional and academic background, the company's new products and important events, the issue they published including papers, patents etc. , and the analysis on the public speech from the CEOs and any other top management staff.

The technical capabilities recognition of all the executives in WorkFusion is higher than management and leadership capabilities. For personal capability, each executive has his or her own strength on an aspect and a few numbers of them do well in all three aspects. Compared with its competitive company, Work Fusion do not show a balance in skill recognition.

The members of the senior management team are generally undergraduates, and most of them majored in job-matching subjects. As a management, the main advantage comes from a rich career background, since most people have more than 20 years of work experience in their respective fields. The frequency of product updated of WORKFUSION is in line with the basic situation of the industry. In recent years, it has also received various commendations and nominations.

No matter numbers of papers, articles, news, patents and events or patents, papers citations, it is not too much but above average level among similar competitors and there are several talents in academic and technology fields. The leadership impact has the medium level according to CB rank.

And then we managed to transform the online videos into text to make analysis on what they said in the speeches and make a judgement on their professional levels according to the text analysis including sentiment analysis etc.

In the future study, first of all, we plan to develop a score system which combine and quantify all the conclusions above, which provides a methodology to evaluate all the start-ups like WorkFusion, score every individual candidate and give a rank of these similar companies by the factors that we did the research on. Second, our plan in the technology part is to combine and pack these researches into one program with a user interface, which will make a product that can be used in every industry combining the philosophy we used in this research in the future. We hope to provide investment managers from private equities with a reference on the venture capital opportunities, to give every institutional and individual investor a snapshot of the industry they are planning to invest in. Third, our results in voice-to-text and text analysis have more information to be mined out. We are planning to utilize the same methods as "Opinion finder" to give a deeper analysis on the what they said in the videos and their Twitters and other posts in social media. It will give a more comprehensive view on the characters of the leaders.

Appendix: Peer View

Name: WU Jianmei ID: 20550001

Mainly collect information of senior managers of WorkFuion from LinkedIn, such as personal skills recognition, and analyze the collected information comprehensively. Do part of the reflection on finance essay.

Name: ZHANG Jian ID:20551079

Python Voice-to-text platform implementation and coding and managed to transform a bunch of speech videos given by executives to text form with the help of Python and Google Could AI and Machine learning platform; Report and slides' finalizing.

HUANG BAOHUI ID: 20549868

Collect executive team member background including academic qualification, major-job matching situation and distribution of industry where the former companies belong; Collect Significant news and product release information.; Reflection on Finance paper; Part of video recording

ZHANG Qianying ID: 20549466

Collect data(numbers of papers, news, articles, patents, projects and events) of Workfusion and its 3 competitors(Thoughtonomy, UiPath and Kryon Systems.) and collect numbers of these data for every current team members of Workfusion, then find every members' Crunchbase ranks and company Crunchbase ranks of 4 companies. Finally, list data of patents and papers citations of Workfusion; Do reflection on an AI essay

YUE Yangshanhui ID: 20551067

Mainly collected videos and speeches of senior executives from YouTube, sent them to my partner, and did analysis on the received transferred texts. For analysis, mainly wrote Python codes to generate word clouds as well as assess professionality, positivity and objectivity of the speeches and speakers by TF-IDF algorithm and Sentiment Analysis. Also ran the codes to compute the results and built a weighting model to evaluate those senior executives from the perspective of their speeches.

Appendix: Figures

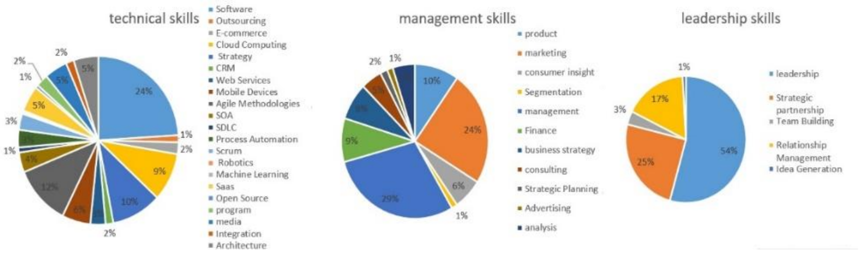


Figure 1: Skill points classification

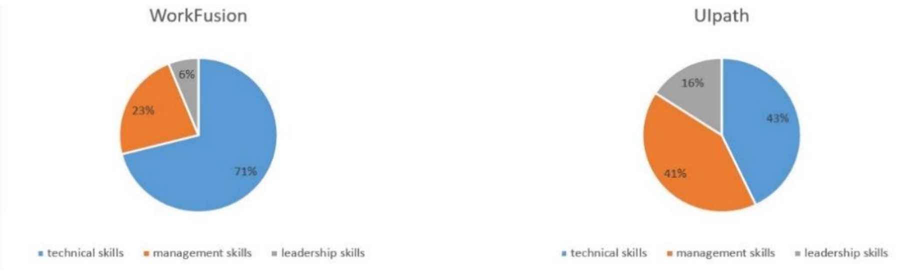


Figure 2: Total skill points comparison

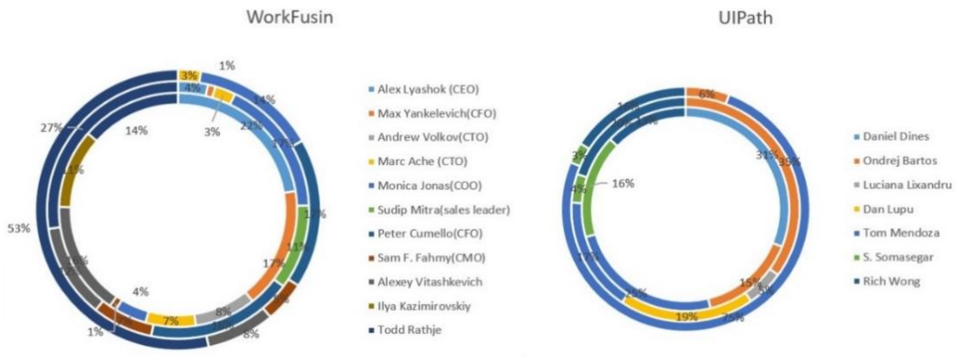


Figure 3: Individual skill points comparison

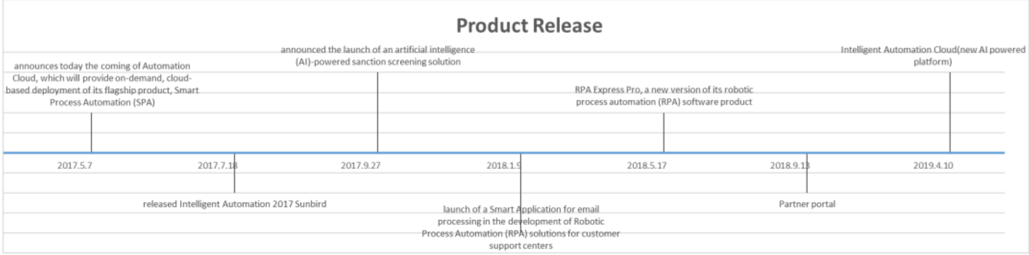


Figure 4: Product Release

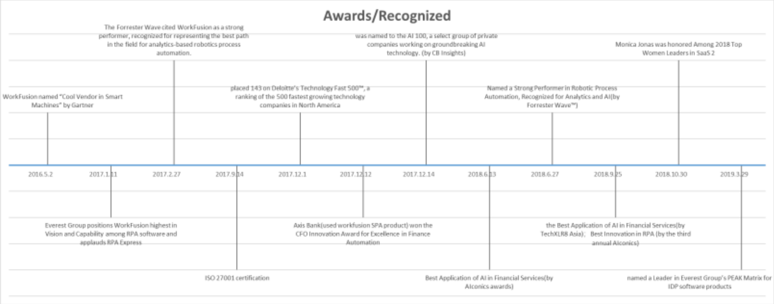


Figure 5: Awards/Recognized

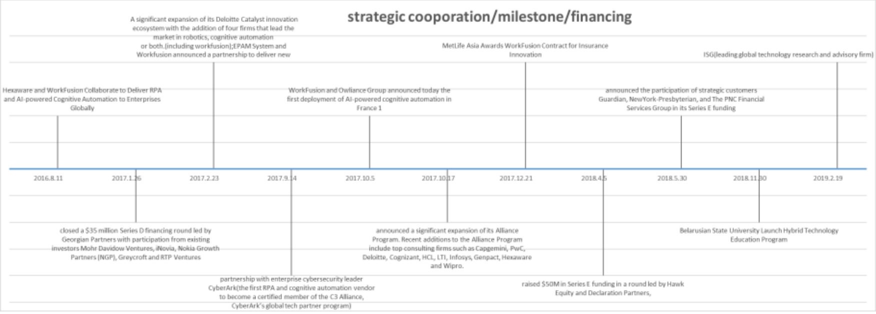

strategic cooporation/milestone/financing
A significant expansion of its Deloitte Catalyst innovation ecosystem with the addition of four firms that lead the market in robotics, cognitive automation or both.(including workfusion);EPAM System and Workfusion announced a partnership to deliver new
MetLife Asia Awards WorkFusion Contract for Insurance Innovation
ISG(leading global technology research and advisory firm)
Hexaware and WorkFusion Collaborate to Deliver RPA and AI-powered Cognitive Automation to Enterprises Globally
WorkFusion and Owliance Group announced today the first deployment of AI-powered cognitive automation in France 1
announced the participation of strategic customers Guardian, NewYork-Presbyterian, and The PNC Financial Services Group in its Series E funding
2016.8.11
2017.1.26
2017.2.23
2017.9.14
2017.10.5
2017.10.17
2017.12.21
2018.4.5
2018.5.30
2018.11.30
2019.2.19
closed a $35 million Series D financing round led by Georgian Partners with participation from existing investors Mohr Davidow Ventures, iNovia, Nokia Growth Partners (NGP), Greycroft and RTP Ventures
announced a significant expansion of its Alliance Program. Recent additions to the Alliance Program include top consulting firms such as Capgemini, PwC, Deloitte, Cognizant, HCL, LTI, Infosys, Genpact, Hexaware and Wipro.
Belarusian State University Launch Hybrid Technology Education Program
partnership with enterprise cybersecurity leader CyberArk(the first RPA and cognitive automation vendor to become a certified member of the C3 Alliance, CyberArk's global tech partner program)
raised $50M in Series E funding in a round led by Hawk Equity and Declaration Partners,

Figure 6: Strategic cooperation/milestone/financing

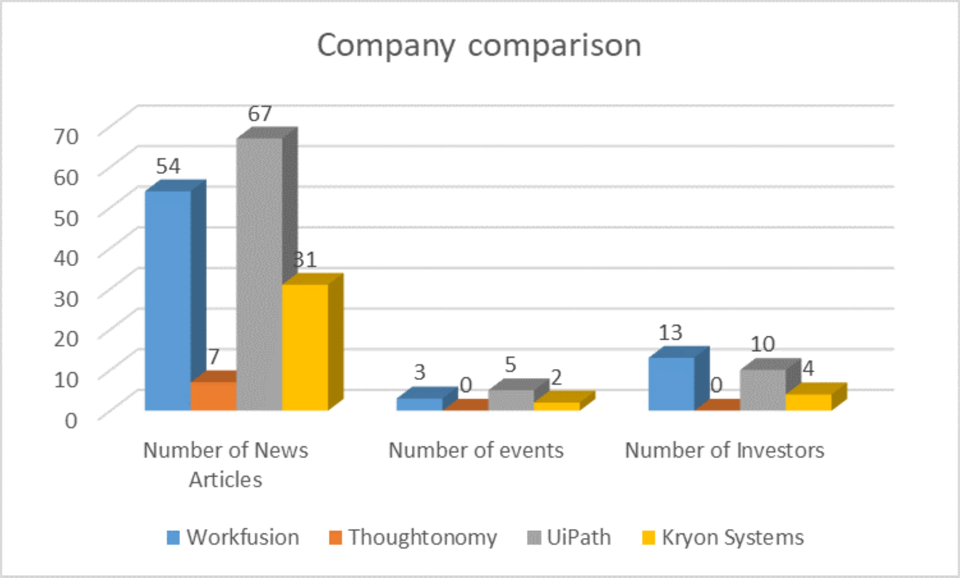


Figure 7: Numbers of News/Articles/events among companies

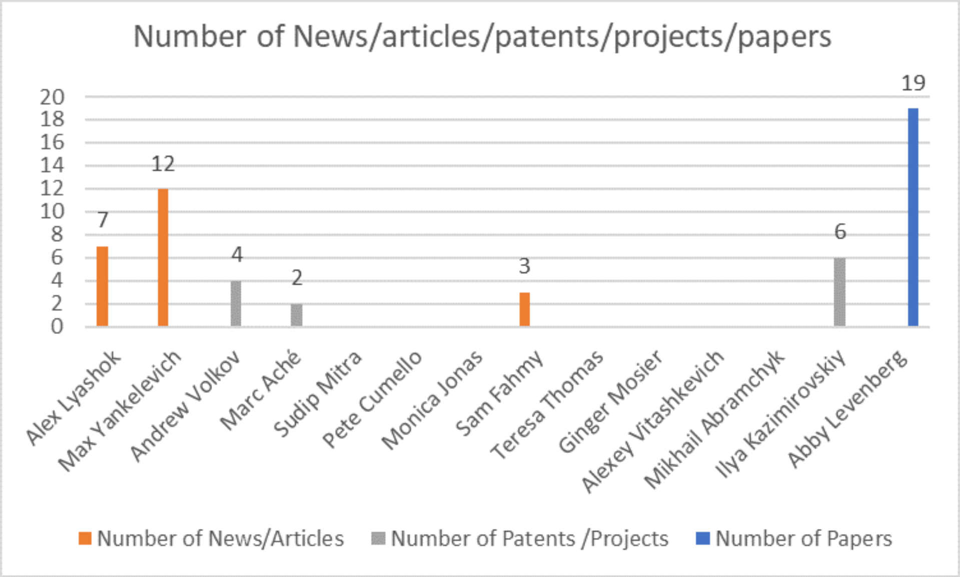


Figure 8: Numbers of News/Articles/events in workfusion

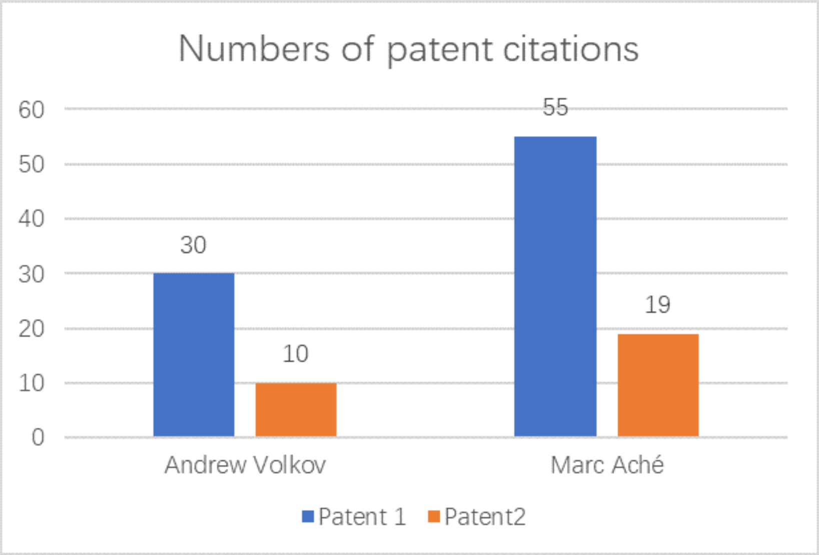


Figure 9: Numbers of patent citations

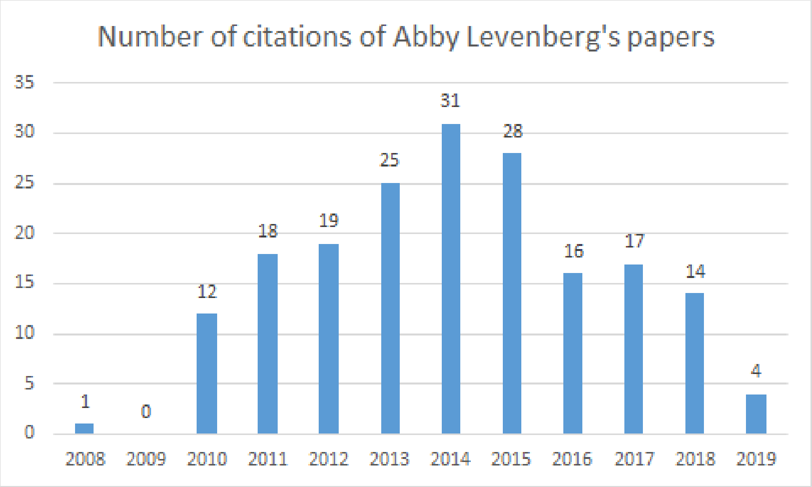


Figure 10: Numbers of citations of Abby Levenberg’s papers

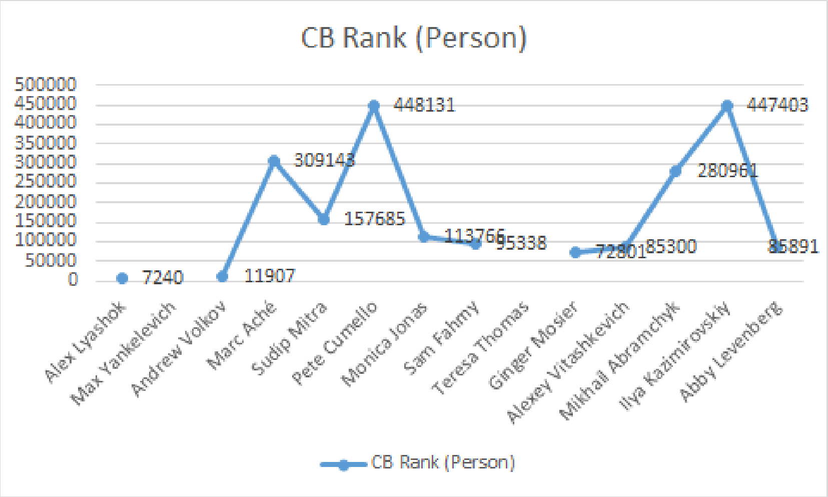


Figure 11: Citations and CB rank

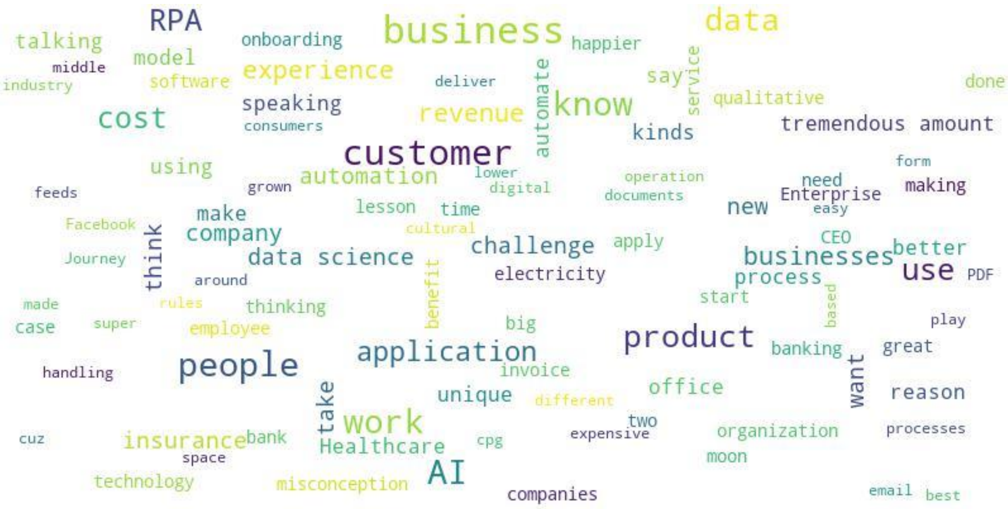


Figure 12: AI In The Here And Now Interview With Adam Devine SVP Marketing WorkFusion Wordcloud

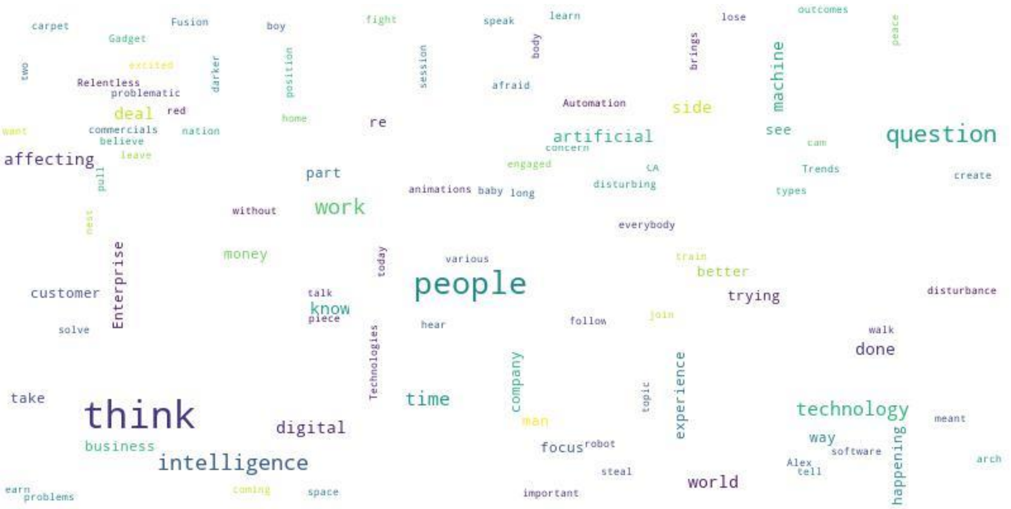


Figure 13: Alex Layshok COO WorkFusion @MeetTheDisruptors NILF 2017h Wordcloud

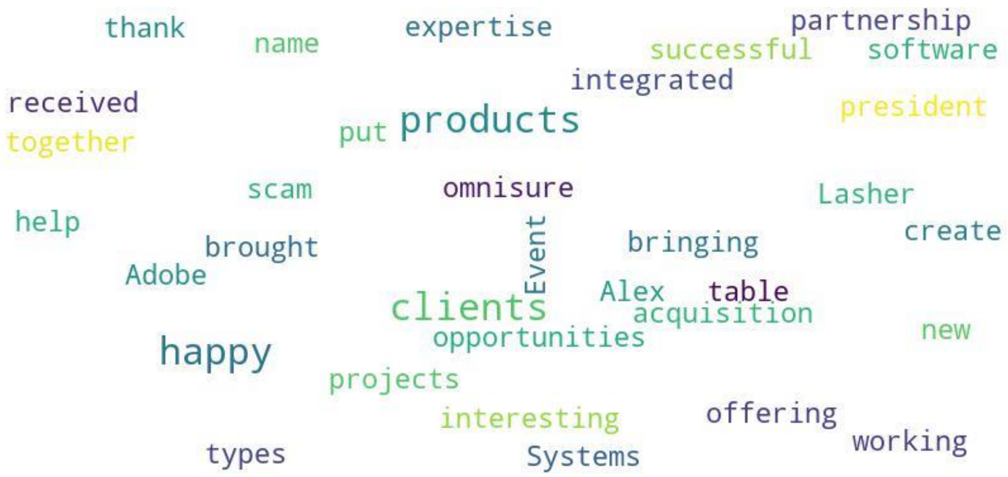


Figure 14: Alex Lyashok from EPAM at DayIgnite about Adobe Wordcloud

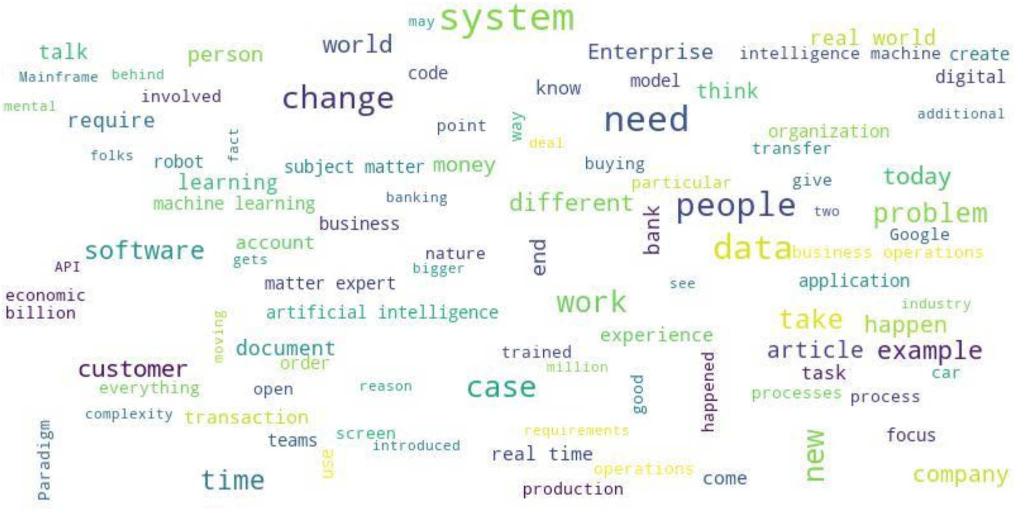


Figure 15: Alex speech-Workfusion Keynote Wordcloud

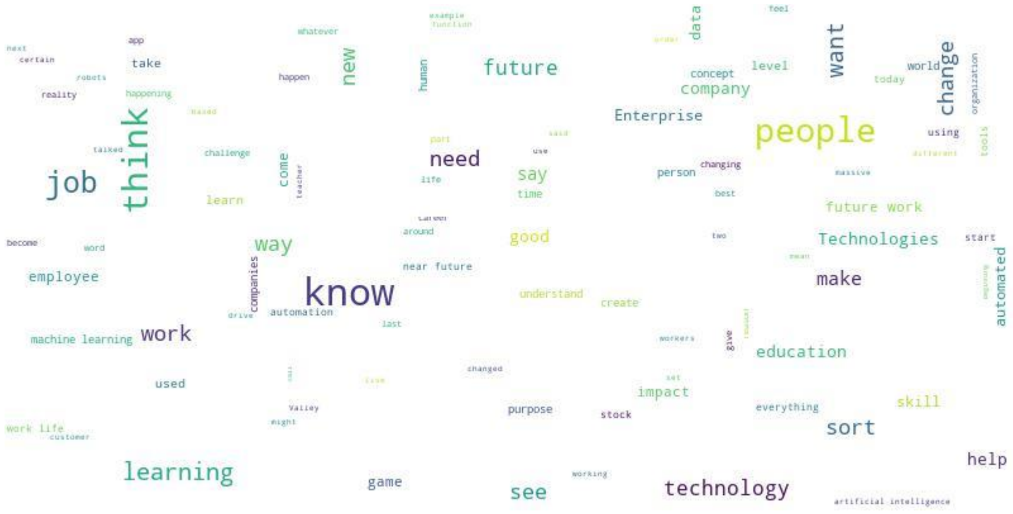


Figure 16: ASU GSV Summit The Here and Now Impact of Emerging Technologies on Work Wordcloud

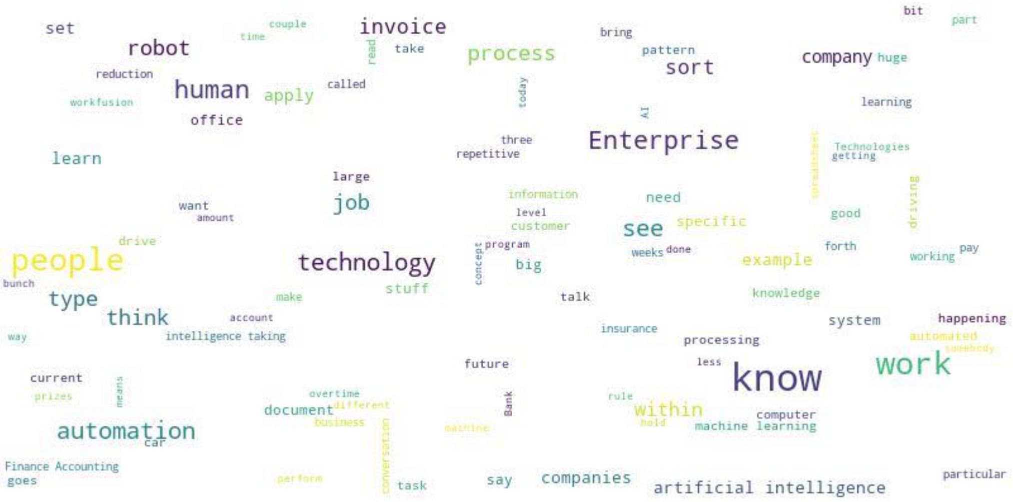


Figure 17: ASU GSV SUMMIT WorkFusion Wordcloud

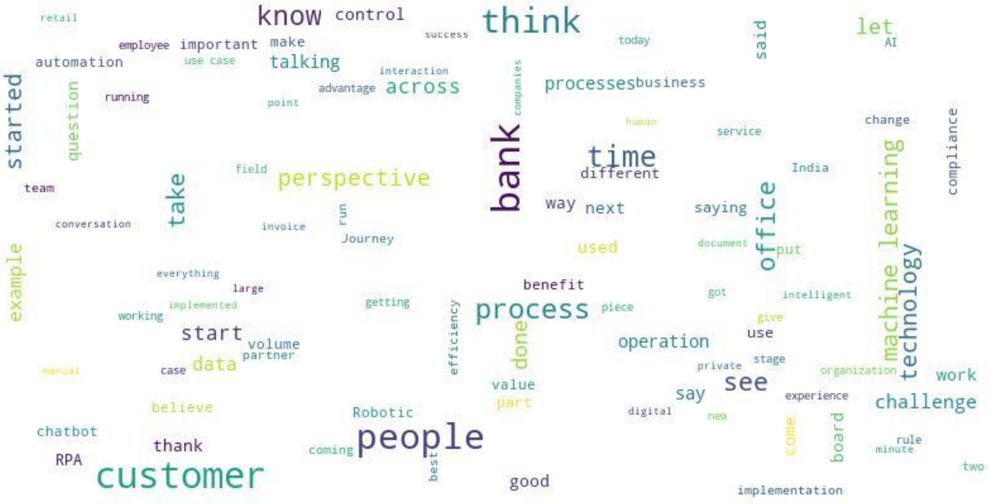


Figure 18: FIBAC 2017 Session Robotics and AI Wordcloud

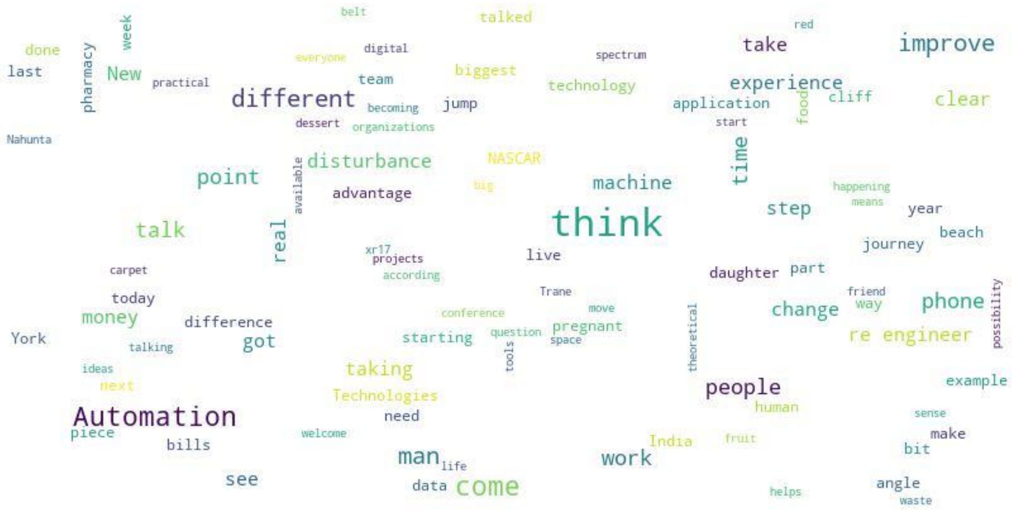


Figure 19: Gautam P Moorjani SVP @ WorkFusion - Cognitive   Robotic Automation for Enterprise   NILF 2017 Wordcloud

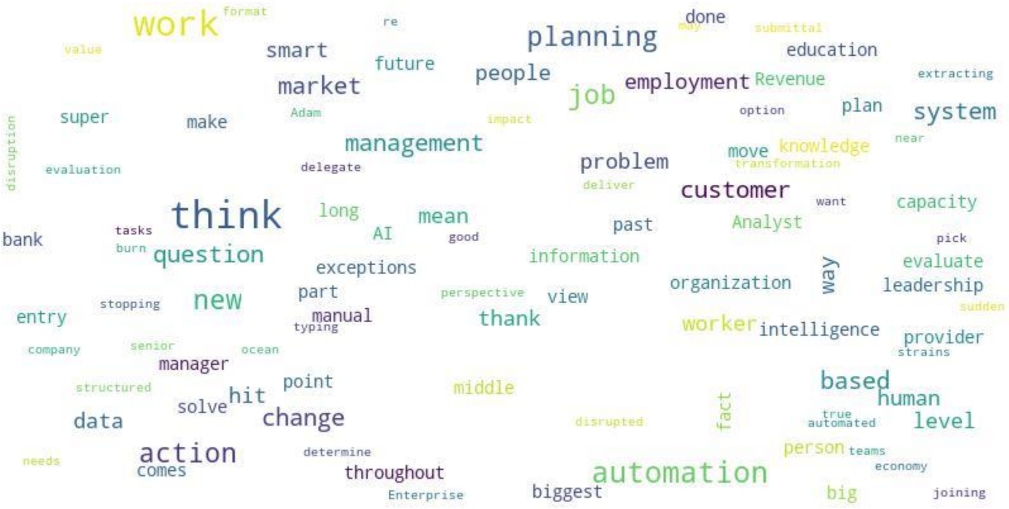


Figure 20: Is AI The Next Big Disruption   Adam Devine VP of Marketing Workfusion NILF 2016 Wordcloud

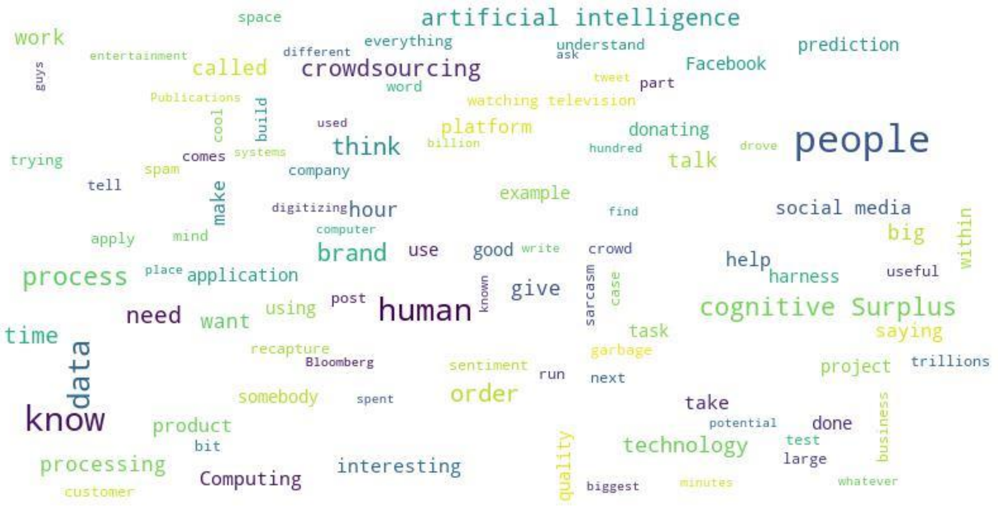


Figure 21: Max Yankelevich CrowdComputing Systems Data Driven NYC 8 Oct 2012 Wordcloud

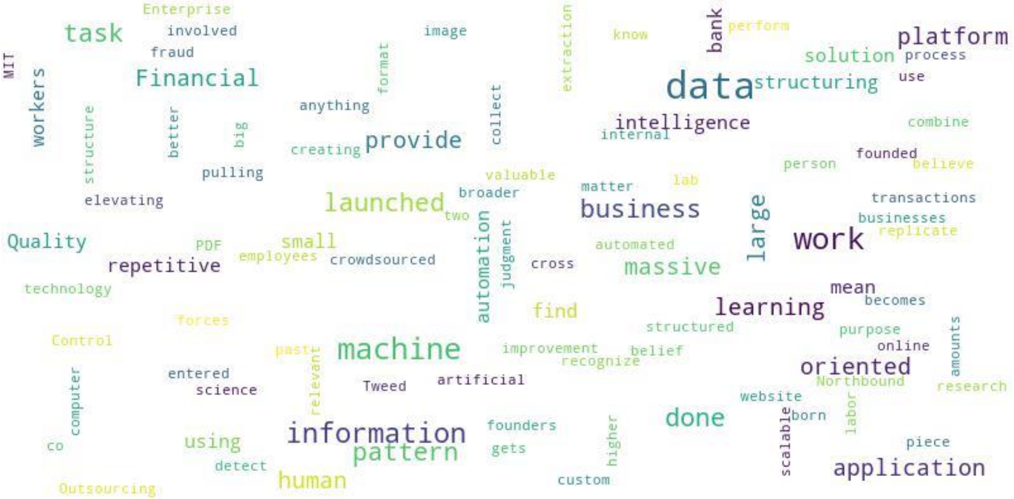
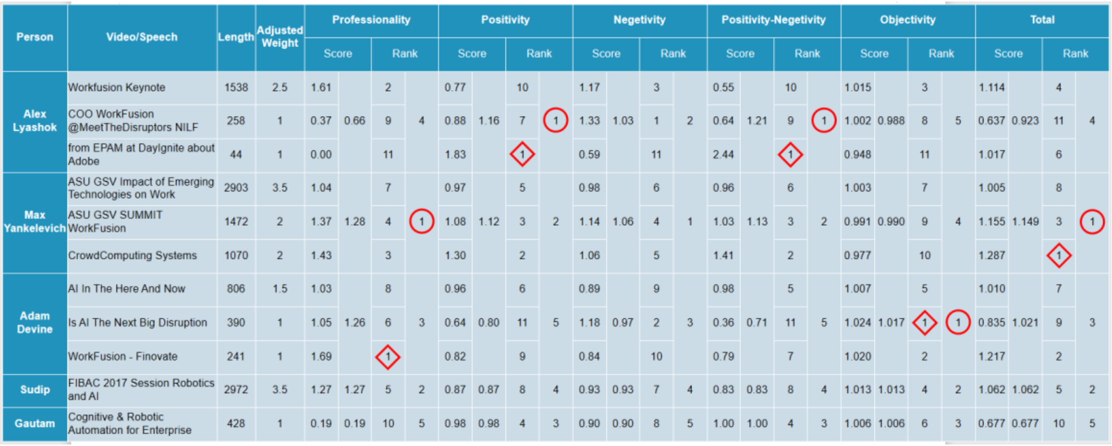


Figure 22: WorkFusion – Finovate Wordcloud



Person	Video/Speech	Length	Adjusted Weight	Professionality				Positivity				Negetivity				Positivity-Negetivity				Objectivity				Total			
				Score		Rank		Score		Rank		Score		Rank		Score		Rank		Score		Rank		Score		Rank	
Alex Lyashok	Workfusion Keynote	1538	2.5	1.61	0.66	2	4	0.77	1.16	10	1	1.17	1.03	3	2	0.55	1.21	10	1	1.015	0.988	3	5	1.114	0.923	4	4
	COO WorkFusion @MeetTheDisruptors NILF	258	1	0.37		9		0.88		7		1.33		1		0.64		9		1.002		8		0.637		11	
	from EPAM at DayIgnite about Adobe	44	1	0.00		11		1.83		1		0.59		11		2.44		1		0.948		11		1.017		6	
Max Yankelevich	ASU GSV Impact of Emerging Technologies on Work	2903	3.5	1.04	1.28	7	1	0.97	1.12	5	2	0.98	1.06	6	1	0.96	1.13	6	2	1.003	0.990	7	4	1.005	1.149	8	1
	ASU GSV SUMMIT WorkFusion	1472	2	1.37		4		1.08		3		1.14		4		1.03		3		0.991		9		1.155		3	
	CrowdComputing Systems	1070	2	1.43		3		1.30		2		1.06		5		1.41		2		0.977		10		1.287		1	
Adam Devine	AI In The Here And Now	806	1.5	1.03	1.26	8	3	0.96	0.80	6	5	0.89	0.97	9	3	0.98	0.71	5	5	1.007	1.017	5	1	1.010	1.021	7	3
	Is AI The Next Big Disruption	390	1	1.05		6		0.64		11		1.18		2		0.36		11		1.024		1		0.835		9	
	WorkFusion - Finovate	241	1	1.69		1		0.82		9		0.84		10		0.79		7		1.020		2		1.217		2	
Sudip	FIBAC 2017 Session Robotics and AI	2972	3.5	1.27	1.27	5	2	0.87	0.87	8	4	0.93	0.93	7	4	0.83	0.83	8	4	1.013	1.013	4	2	1.062	1.062	5	2
Gautam	Cognitive & Robotic Automation for Enterprise	428	1	0.19	0.19	10	5	0.98	0.98	4	3	0.90	0.90	8	5	1.00	1.00	4	3	1.006	1.006	6	3	0.677	0.677	10	5

Figure 23: Text Analysis Rankings

Appendix: Charts

Table 1: Pre-Industry

Software	EPAM, Freedom OSS, IBM, Bungo, Yahoo, Hewlett Packard
Media	Viacom, Pressplay, Washingtonpost. Newsweek Interactive, PBS, Glam Media, Current TV
Music	Universal Music, Napster
Management Consulting	Accenture, WNS, Cvent
Marketing Communication	Text100, Tacoda System
Government Agency	NASA
Data Analysis	Dun & Bradstreet
Clothing	LeviStrauss & Co

Table 2: CB rank among companies

company	CB Rank
Workfusion	1004
Thoughtonomy	56359
UiPath	671
Kryon Systems	1904

Appendix: Coding

Speech to text:

```
# -*- coding: utf-8 -*-
"""
Created on Sun Apr 28 12:31:08 2019
@author: K.
"""
import os
import io
from google.cloud import speech
from google.cloud.speech import enums
from google.cloud.speech import types
import pandas as pd
```

```
os.environ['GOOGLE_APPLICATION_CREDENTIALS'] =
"C:/Users/asus/Documents/My First Project-d141094bd68d.json"
client = speech.SpeechClient()

# Loads the audio into memory



audio = types.RecognitionAudio(uri='gs://filesai/ASU GSV SUMMIT
WorkFusion.flac')
config = types.RecognitionConfig(
    encoding=enums.RecognitionConfig.AudioEncoding.FLAC,
        sample_rate_hertz=16000,
        language_code='en-US')

# Detects speech in the audio file
operation = client.long_running_recognize(config, audio)
print('Waiting for operation to complete...')
response = operation.result(timeout=9000)

trans = []
conf = []
txt = ''
for result in response.results:
    # The first alternative is the most likely one for this portion.
    print(u'Tranlate: {}'.format(result.alternatives[0].transcript))
    print(u'Confidence: {}'.format(result.alternatives[0].confidence))
    trans.append(u'{}'.format(result.alternatives[0].transcript))
    conf.append(u'{}'.format(result.alternatives[0].confidence))
    txt = txt + u'{}'.format(result.alternatives[0].transcript) +'/'
text_file = open("ASU GSV SUMMIT WorkFusion.txt", "w")
text_file.write(txt)
text_file.close()
data = pd.concat([pd.DataFrame(trans),pd.DataFrame(conf)],axis =1)
```

```
data.to_excel('ASU GSV SUMMIT WorkFusion.xlsx')
```

Sentiment Analysis:

```
from nltk.classify import NaiveBayesClassifier
from nltk.corpus import subjectivity
from nltk.sentiment import SentimentAnalyzer
from nltk.sentiment.util import *
import os
from os import path
from nltk.corpus import sentiwordnet as swn
import pandas as pd

def isSymbol(inputString):
    return bool(re.match(r'[^\w]', inputString))
from nltk.corpus import stopwords
stop = stopwords.words('english')
from nltk.stem import WordNetLemmatizer
wordnet_lemmatizer = WordNetLemmatizer()

def check(word):

    word= word.lower()
    if word in stop:
        return False
    elif isSymbol(word):
        return False
    else:
        return True

def preprocessing(sen):
    res = []
    for index in range(len(sen)):
        if check(sen[index]):
            res.append(wordnet_lemmatizer.lemmatize(sen[index]))
    return res

d = path.dirname(__file__) if "__file__" in locals() else os.getcwd()

# Read the whole text.
file_import_url_list=['E:\\HKUST_courses\\AI\\project\\output\\AI  In
The Here And Now Interview With Adam Devine SVP Marketing
WorkFusion.txt',
                    'E:\\HKUST_courses\\AI\\project\\output\\Alex
Layshok COO WorkFusion @MeetTheDisruptors NILF 2017h.txt',
                    'E:\\HKUST_courses\\AI\\project\\output\\Alex
Lyashok from EPAM at DayIgnite about Adobe.txt',
                    'E:\\HKUST_courses\\AI\\project\\output\\Alex
speech.txt',
                    'E:\\HKUST_courses\\AI\\project\\output\\ASU  GSV
Summit The Here and Now Impact of Emerging Technologies on Work.txt',
                    'E:\\HKUST_courses\\AI\\project\\output\\FIBAC
2017 Session Robotics and AI.txt',
                    'E:\\HKUST_courses\\AI\\project\\output\\Gautam P
Moorjani SVP @ WorkFusion - Cognitive  Robotic Automation for
Enterprise  NILF 2017.txt',
                    'E:\\HKUST_courses\\AI\\project\\output\\Is  AI
The Next Big Disruption  Adam Devine VP of Marketing Workfusion  NILF
2016.txt',
                    'E:\\HKUST_courses\\AI\\project\\output\\Max
Yankelevich CrowdComputing Systems  Data Driven NYC 8  Oct 2012.txt',

'E:\\HKUST_courses\\AI\\project\\output\\WorkFusion - Finovate.txt',
                    'E:\\HKUST_courses\\AI\\project\\output\\ASU  GSV
SUMMIT WorkFusion.txt']
file_export_url_list=['E:\\HKUST_courses\\AI\\project\\output\\rank_
by_positivity.xlsx',

'E:\\HKUST_courses\\AI\\project\\output\\rank_by_negativity.xlsx',
'E:\\HKUST_courses\\AI\\project\\output\\rank_by_objectivity.xlsx',

'E:\\HKUST_courses\\AI\\project\\output\\rank_by_pos_neg_difference.
xlsx',

'E:\\HKUST_courses\\AI\\project\\output\\rank_by_words_number.xlsx']
pos_result={}
neg_result={}
obj_result={}
pos_neg_diff={}
words_num={}

for i in range(len(file_import_url_list)):

filename=file_import_url_list[i].strip('E:\\HKUST_courses\\AI\\proje
ct\\output\\').strip('.txt')
    text = open(path.join(d, file_import_url_list[i])).read()
    word_temp=nltk.word_tokenize(text)
    words=preprocessing(word_temp)

    pos_score=0
    neg_score=0
    obj_score=0
    count=0

    for word in words:
        word_transfer=list(swn.senti_synsets(word))
        if len(word_transfer)!=0:
            pos_score=pos_score+word_transfer[0].pos_score()
            neg_score=neg_score+word_transfer[0].neg_score()
            obj_score=obj_score+word_transfer[0].obj_score()
            count+=1
    print(count)
    pos_result[filename]=pos_score/count
    neg_result[filename]=neg_score/count

    obj_result[filename]=obj_score/count

    pos_neg_diff[filename]=(pos_score-neg_score)/count

    words_num[filename]=count

    pos_result_rank=sorted(pos_result.items(),key=lambda
k:k[1],reverse=True)

    neg_result_rank=sorted(neg_result.items(),key=lambda
k:k[1],reverse=True)

    obj_result_rank=sorted(obj_result.items(),key=lambda
k:k[1],reverse=True)

    pos_neg_diff_rank=sorted(pos_neg_diff.items(),key=lambda
k:k[1],reverse=True)

    words_num_rank=sorted(words_num.items(),key=lambda
k:k[1],reverse=True)



    pd.DataFrame(pos_result_rank).to_excel(file_export_url_list[0])

    pd.DataFrame(neg_result_rank).to_excel(file_export_url_list[1])

    pd.DataFrame(obj_result_rank).to_excel(file_export_url_list[2])

    pd.DataFrame(pos_neg_diff_rank).to_excel(file_export_url_list[3])

    pd.DataFrame(words_num_rank).to_excel(file_export_url_list[4])
```

Text Analysis

```
import re

import nltk

from functools import reduce

from __future__ import division

import math

import os

from os import path

import nltk

import pandas as pd
```

```
def isSymbol(inputString):
    return bool(re.match(r'[^\w]', inputString))
def hasNumbers(inputString):
    return bool(re.search(r'\d', inputString))
from nltk.corpus import stopwords
stop = stopwords.words('english')
from nltk.stem import WordNetLemmatizer
wordnet_lemmatizer = WordNetLemmatizer()

def check(word):

    word= word.lower()
    if word in stop:
        return False
    elif hasNumbers(word) or isSymbol(word):
        return False
    else:
        return True



def preprocessing(sen):
    res = []
    for index in range(len(sen)):
        if check(sen[index]):
            res.append(wordnet_lemmatizer.lemmatize(sen[index]))
    return res

def TF_IDF_Compute(file_import_url_list,file_export_url,words):

    InputFormatList=['utf-8']
    OutputFormatList=['utf-8']
    pattern="full"
    n=10
    ruler_list=[]
    result_file_num=50
    out_to_file=True

    word_in_allfiles_stat={}
    word_in_afile_stat={}
    files_num=len(file_import_url_list)

    for index in range(files_num):
        d = path.dirname(__file__) if "__file__" in locals() else
os.getcwd()
        data = open(path.join(d, file_import_url_list[index])).read()


        word_temp=nltk.word_tokenize(data)  #key words of a file
        data_temp=preprocessing(word_temp)


file_name=file_import_url_list[index].strip('E:\\HKUST_courses\\AI\\
project\\output').strip('.txt')
        data_temp_len=len(data_temp)


        for word in words:
            if word in data_temp:

                if not word_in_allfiles_stat.__contains__(word):
                    word_in_allfiles_stat[word]=1
                else:
                    word_in_allfiles_stat[word]+=1
                if not word_in_afile_stat.__contains__(file_name):
                    word_in_afile_stat[file_name]={}
                if not
word_in_afile_stat[file_name].__contains__(word):
                    word_in_afile_stat[file_name][word]=[]
word_in_afile_stat[file_name][word].append(data_temp.count(word))
word_in_afile_stat[file_name][word].append(data_temp_len)

    if (word_in_afile_stat) and (word_in_allfiles_stat) and
(files_num !=0):
        TF_IDF_result={}
        for filename in word_in_afile_stat.keys():
            TF_IDF_result[filename]={}
            for word in word_in_afile_stat[filename].keys():
                word_n=word_in_afile_stat[filename][word][0]
                word_sum=word_in_afile_stat[filename][word][1]
                with_word_sum=word_in_allfiles_stat[word]
TF_IDF_result[filename][word]=((word_n/word_sum))*(math.log10(files_
num/with_word_sum))

        TF_IDF_total={}
        for filename in TF_IDF_result.keys():
            TF_IDF_total[filename]=reduce(lambda
x,y:x+y,TF_IDF_result[filename].values())
        result_temp=[]
        result_temp=sorted(TF_IDF_total.items(),key=lambda
x:x[1],reverse=True)

        k=result_file_num
        result=[]
        for item in result_temp:
            if k!=0:

                result.append(item[0]+' '+str(item[1]))

                k-=1

            else:

                break


    else:

        result=["None"]


    # print(TF_IDF_result)

    for index in range(files_num):

file_name=file_import_url_list[index].strip('E:\\HKUST_courses\\AI\\project\\output\\').strip('.txt')

file_export_url_temp=file_import_url_list[index].strip('.txt')+'_result.xlsx'

        output_all=pd.DataFrame

        if file_name in TF_IDF_result.keys():

            output_excel=pd.DataFrame([TF_IDF_result[file_name]])

            output_excel.to_excel(file_export_url_temp)



    if out_to_file:


        pd.DataFrame(result_temp).to_excel(file_export_url)


    else:

        return result
```

Word Cloud

```
%matplotlib inline
import re
import os
from os import path
from wordcloud import WordCloud,STOPWORDS
import nltk
import matplotlib.pyplot as plt

# get data directory (using getcwd() is needed to support running
example in generated IPython notebook)
d = path.dirname(__file__) if "__file__" in locals() else os.getcwd()

# Read the whole text.
file_import_url_list=['E:\\HKUST_courses\\AI\\project\\output\\AI In
The Here And Now Interview With Adam Devine SVP Marketing
WorkFusion.txt',
                    'E:\\HKUST_courses\\AI\\project\\output\\Alex
Layshok COO WorkFusion @MeetTheDisruptors NILF 2017h.txt',
                    'E:\\HKUST_courses\\AI\\project\\output\\Alex
Lyashok from EPAM at DayIgnite about Adobe.txt',
                    'E:\\HKUST_courses\\AI\\project\\output\\Alex
speech.txt',
                    'E:\\HKUST_courses\\AI\\project\\output\\ASU  GSV
Summit The Here and Now Impact of Emerging Technologies on Work.txt',
                    'E:\\HKUST_courses\\AI\\project\\output\\FIBAC
2017 Session Robotics and AI.txt',
                    'E:\\HKUST_courses\\AI\\project\\output\\Gautam P
Moorjani SVP @ WorkFusion - Cognitive  Robotic Automation for
Enterprise  NILF 2017.txt',
                    'E:\\HKUST_courses\\AI\\project\\output\\Is  AI
The Next Big Disruption  Adam Devine VP of Marketing Workfusion  NILF
2016.txt',
                    'E:\\HKUST_courses\\AI\\project\\output\\Max
Yankelevich CrowdComputing Systems  Data Driven NYC 8  Oct 2012.txt',

'E:\\HKUST_courses\\AI\\project\\output\\WorkFusion - Finovate.txt',
                    'E:\\HKUST_courses\\AI\\project\\output\\ASU  GSV
SUMMIT WorkFusion.txt']


def preprocessing(text):
    words = nltk.word_tokenize(text)
    tagged = nltk.pos_tag(words)
    res = []
    for index in range(len(words)):
        if tagged[index][1].find('RB')!=-1:
            res.append(wordnet_lemmatizer.lemmatize(words[index]))
    return res

for i in range(len(file_import_url_list)):
    text = open(path.join(d, file_import_url_list[i])).read()

# Generate a word cloud image
    wordcloud = WordCloud().generate(text)

# Display the generated image:
# the matplotlib way:

    plt.imshow(wordcloud, interpolation='bilinear')
    plt.axis("off")

# lower max_font_size
    wordcloud = WordCloud(max_font_size=30).generate(text)
    plt.figure()
    plt.imshow(wordcloud, interpolation="bilinear")
    plt.axis("off")
    plt.show()

# add stopwords
    stopwords = set(STOPWORDS)
    stopadds = preprocessing(text)
    stopadds2                                                      =
['thing','will','kind','something','every','one','lot','much','able'
,'vice','go','many','another','going',

'still','right','year','month','day','many','second','first','term',
'even','looking','little','things','years',
               'days','months','terms','look','us']
    for j in range(len(stopadds)):
        stopwords.add(stopadds[j])
    for j2 in range(len(stopadds2)):
        stopwords.add(stopadds2[j2])



    wordcloud                                                      =
WordCloud(width=800,height=400,stopwords=stopwords,background_color=
'white',
                        max_font_size=30,max_words=100).generate(text)
    plt.figure()
    plt.imshow(wordcloud, interpolation="bilinear")
    plt.axis("off")
    plt.show()


file_export_url=file_import_url_list[i].strip('.txt')+'_wordcloud_2.
jpg'
    wordcloud.to_file(file_export_url)
```