
데이터 분석보고서

2021. 2

1. 분석목적

- 산업통상자원부 메신저 데이터 분석

2. 분석개요

- 분석과제 : 산업통상자원부 내부 메신저 분석
- 분석수행 : 산업통상자원부 내부 메신저 DB에서 데이터를 추출하여 자체 분석 수행
- 소요예산 : 비예산
- 수행기간 : 2021.02.01. ~ 02.16.
- 활용데이터
 - 메신저 DB에서 추출한 데이터 (2021.01.15 ~ 2021.01.21)
 - 메신저 내용 및 부서 직원 정보 데이터
- 주요내용
 - 데이터 구성
 - 부서 그룹화
 - 메신저를 많이 주고받은 부서(정보관리담당관 기준)
 - 단어 유사도
 - 분석 결과 및 추가사항

Ⅱ 분석결과

1. 내부 메신저 데이터 분석

□ 데이터 구성

○ 메신저 내용 데이터(2021.01.15 ~ 2021.01.21.)

- 81,706건의 데이터와 11건의 칼럼으로 구성
- 칼럼은 메신저 내용, 보내는 사람 이름, 보내는 날짜 및 시간, 받는 사람 등으로 구성

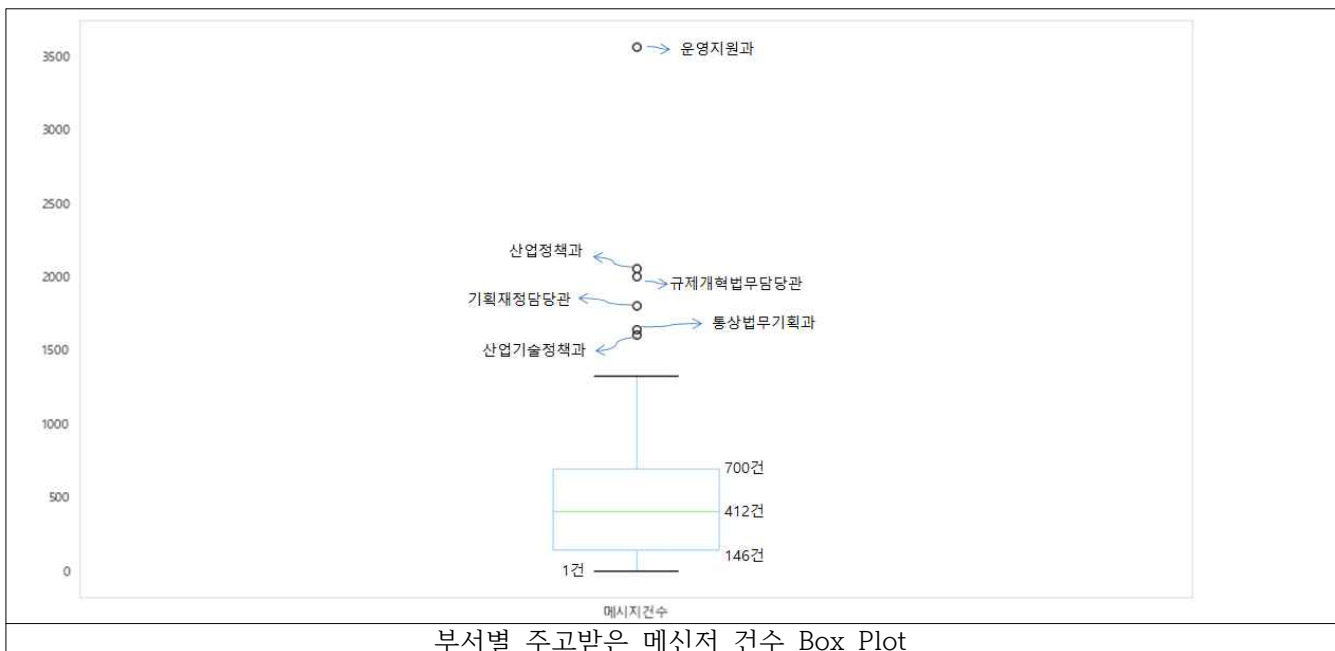
○ 직원 정보 데이터

- 2,255건의 데이터와 6개의 칼럼으로 구성
- 칼럼은 직원이름, 계급, 부서, 담당업무 등으로 구성

□ 부서 그룹화

○ 부서별로 그룹화하여 메신저를 분석 할 필요가 있다고 판단

- 부서의 개수가 185개가 되므로 Box Plot으로 표현



- 최솟값은 1건이며 하위 25%에 해당하는 부서의 주고받은 메신저 건수는 146건, 하위 50%에 해당하는 부서는 412건, 하위 75%에 해당하는 부서는 700건이다.
- 이상치 값들이 존재하는데 운영지원과, 산업정책과, 규제개혁법무담당관 등이 해당된다. 그 중 운영지원과의 경우 3,500건이 넘어가는 부서이고 메신저 활용이 두드러진다.

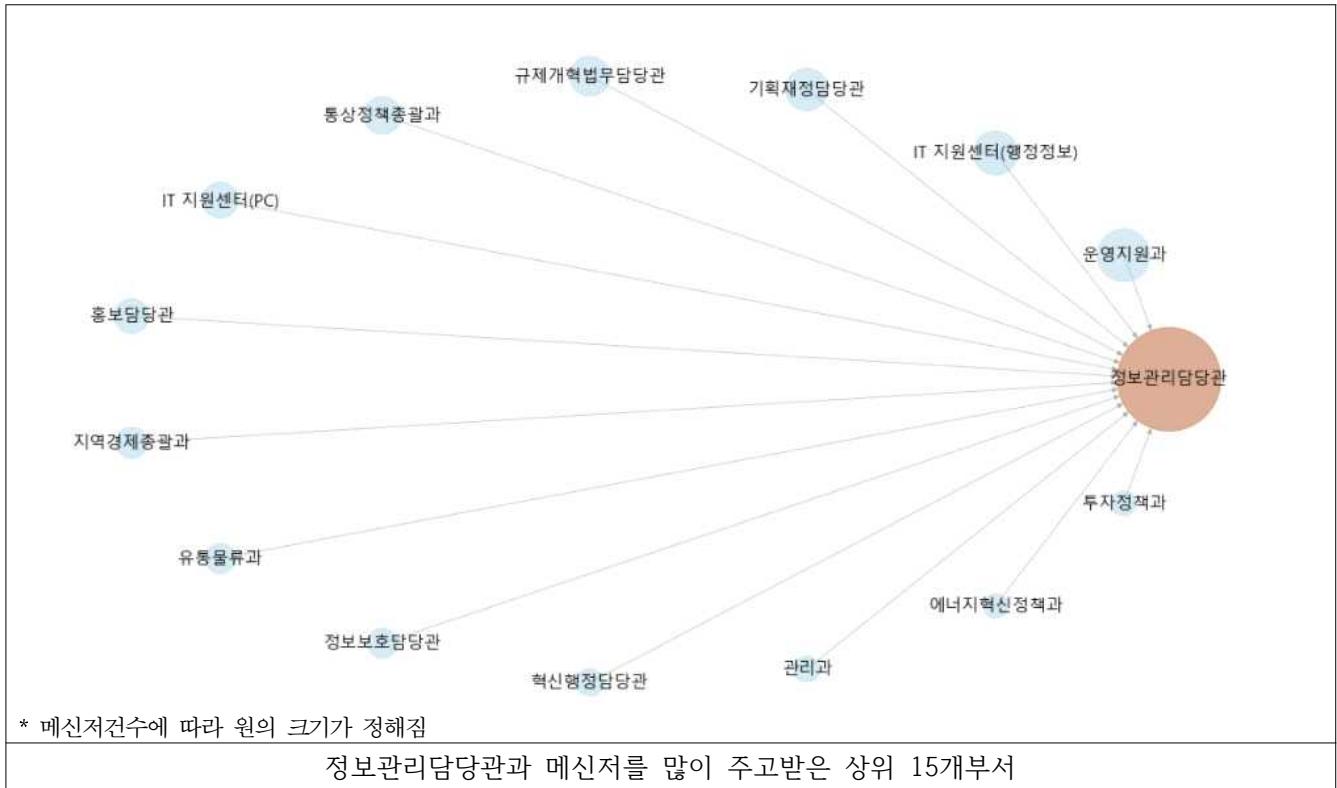
□ 메신저를 많이 주고받은 부서(정보관리담당관 기준)

○ 해당 부서가 어떤 부서에게 메신저를 주고받는지 판단 필요

user_group_name		recv_g	개수
0	IT 지원센터(PC)	IT 지원센터(PC)	18
24	IT 지원센터(국표원)	국내대책과	22
45	IT 지원센터(행정정보)	IT 지원센터(행정정보)	59
73	가스산업과	가스산업과	119
123	감사담당관	감사담당관	215
...
7548	홍보담당관	홍보담당관	384
7634	홍보소통과	홍보소통과	63
7658	화학산업팀	화학산업팀	56
7703	활용촉진팀	활용촉진팀	121
7722	휴직 자유무역협정정책기획과		58
해당부서가 다른부서에게 주고받은 메신저수			

- 칼럼 user_group_name이 기준이 되는 부서이고 recv_g가 메신저를 주고받는 대상이 되는 부서이다.
- 부서가 많은 관계로 가장 많이 주고받는 부서들만 뽑고 정보관리담당관을 예로 들어 분석을 해보겠다.

○ 정보관리담당관과 메시지를 많이 주고받은 상위 15개부서



- 정보관리담당관과 메시지를 많이 주고받은 부서는 운영지원과, IT지원센터(행정정보), 기획재정담당관, 규제개혁법무담당관 등이 있다.
- 정보관리담당관은 부서 특성상 IT관련부서와 소통을 많이 해야 하므로 IT 관련부서와 메시지를 많이 주고받는 것으로 판단된다.
- 정보관리담당관이 주고받을 때 어떤 내용을 주고받는지 알아보겠다.

○ 메신저 내용 전처리

- 메신저내용에 불용어가 많고 답장할 때 “RE:” 가 붙어서 답장이 되므로 텍스트 전처리 작업이 필요

msg_key			reply_key	msg_subject
0	61673BFBABF1D24215BF4594D584D113		NaN	출근보고 드립니다.
1	E4E57E62DE9E35454C1A4ABF62CAA8AE FA99BC89CF9B7ECEBCFE071A60F584AD			Re: 안녕하세요 재우팀 정보입니다.
2	2F028CA2872D0377DBB56A8E721FEDAD		NaN	성공하셨습니다.
3	BE7ECDAABB4FB5A9D42DBAB087680433 6A15B1158E339F7ADB86BCAD558EA37A			Re: Re: Re: 안녕하세요
4	6E524CFD18210F5678BF08DD738125EB ABA7B3A5F18D1904E1CB41B7A18E7AD2			Re: Re: 귀과에 1월 14일자 전문(비밀, 대외비)가 있습니다.

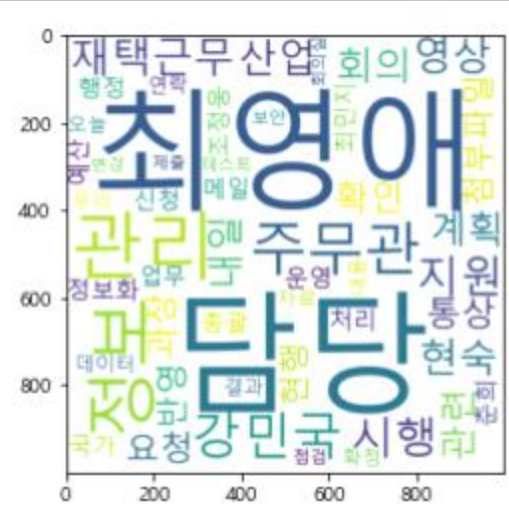
msg_subject	user_name	msg_send_date	msg_rcv_name	msg_srh_content
0	출근보고 드립니다	이승권	20210115000001	활영래 출근보고 드립니다
1	안녕하세요 재우팀 정보입니다	박진석	20210115000012	정호민 안녕하세요공제란에 전역공제원화로 되어 있는데로 해야 하나요 보낸사람 정호민...
2	성공하셨습니다	진상연	20210115000020	유수열 성공하셨습니다
3	안녕하세요	고현	20210115000022	배진석 네 과장님알겠습니다 현재 장관님께 영퇴서 제출하신 상황이고 이후 신원 조회등 절차가

전처리 전 메신저 내용	전처리 후 메신저 내용
--------------	--------------

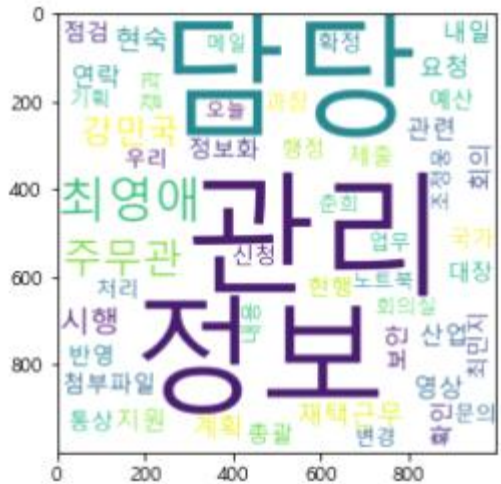
- 메신저 전처리 후 불용어 제거 및 "RE:" 가 제거되었다.

○ 정보관리담당관이 보내는 메신저와 받는 메신저를 분리해 분석


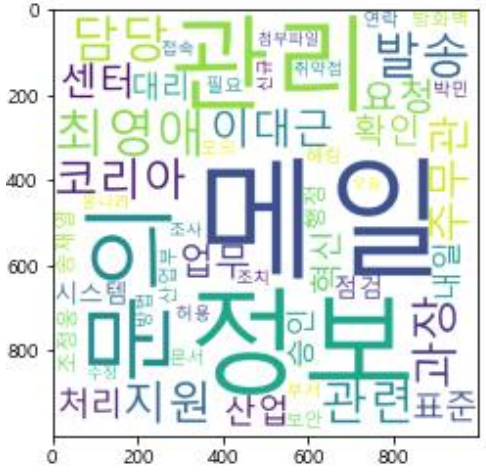
- 정보관리담당관에서 보내는 메시지와 받는 메시지 따로 분리하고 전처리한 메신저 내용들을 명사단위로 추출 후 명사별로 많이 언급된 단어 추출

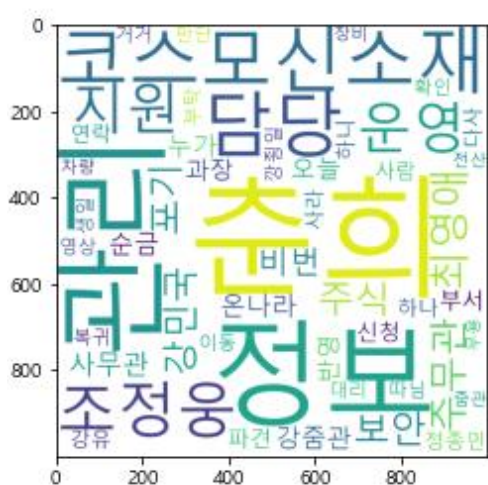
<p>('담당', 700) ('최영애', 683) ('관리', 658) ('정보', 637) ('주무관', 372) ('강민국', 287) ('시행', 213) ('지원', 207) ('재택근무', 186) ('영상', 176) ('현숙', 171) ('산업', 168) ('회의', 160) ('내일', 157) ('계획', 154)</p>	
정보관리담당관이 보내는 메시지에서 언급된 단어	언급된 단어의 개수기반 워드클라우드

- 정보관리담당관이 보내는 메시지에서 가장 많이 사용하는 단어는 “담당”이고 그 다음으로는 “최영애”이다.

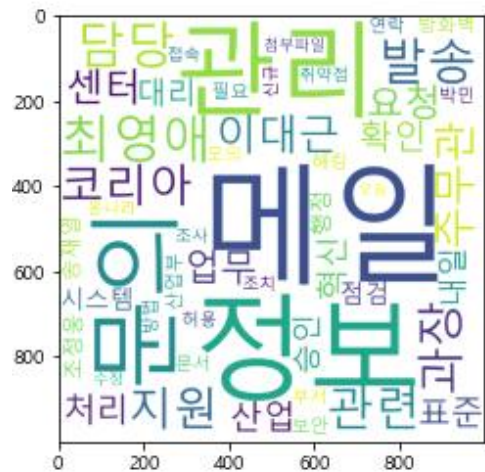
<p>('관리', 757) ('담당', 756) ('정보', 704) ('최영애', 552) ('주무관', 366) ('강민국', 237) ('시행', 193) ('현숙', 161) ('재택근무', 156) ('영상', 143) ('계획', 141) ('요청', 139) ('지원', 139) ('관련', 136) ('연락', 135)</p>	
<p>정보관리담당관이 받는 메시지에서 언급된 단어</p>	<p>언급된 단어의 개수기반 워드클라우드</p>

- 정보관리담당관이 받는 메시지에서 가장 많이 사용하는 단어는 “관리”이고 그 다음으로는 “담당”, “정보” 순이다.
- “관리”, “담당”, “정보”는 정보관리담당관을 지칭할 때 많이 쓰는 단어들이므로 이들을 제외하고 보면 “최영애”라는 단어가 가장 상단에 올라가게 된다. 메시지를 보낼 때나 받을 때 최영애 주무관님이 가장 많이 언급되는 것을 알 수 있다.
- 보낼 때와 받을 때 단어의 유형을 보면 드러나는 특성이 크게 없어 보인다.

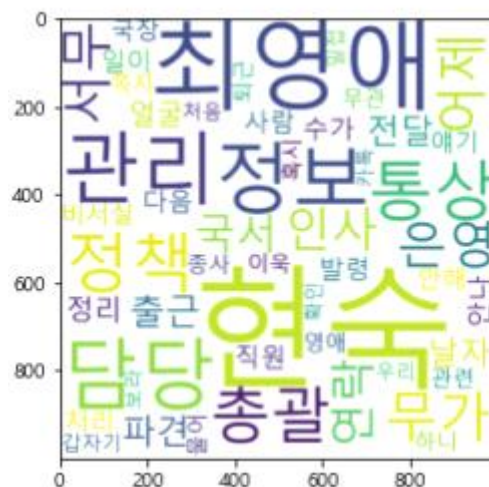
	
<p>정보관리담당관과 운영지원과</p>	<p>정보관리담당관과 IT 지원센터(행정정보)</p>



정보관리담당관과 기획재정담당관



정보관리담당관과 규제개혁법무담당관



정보관리담당관과 통상정책총괄과

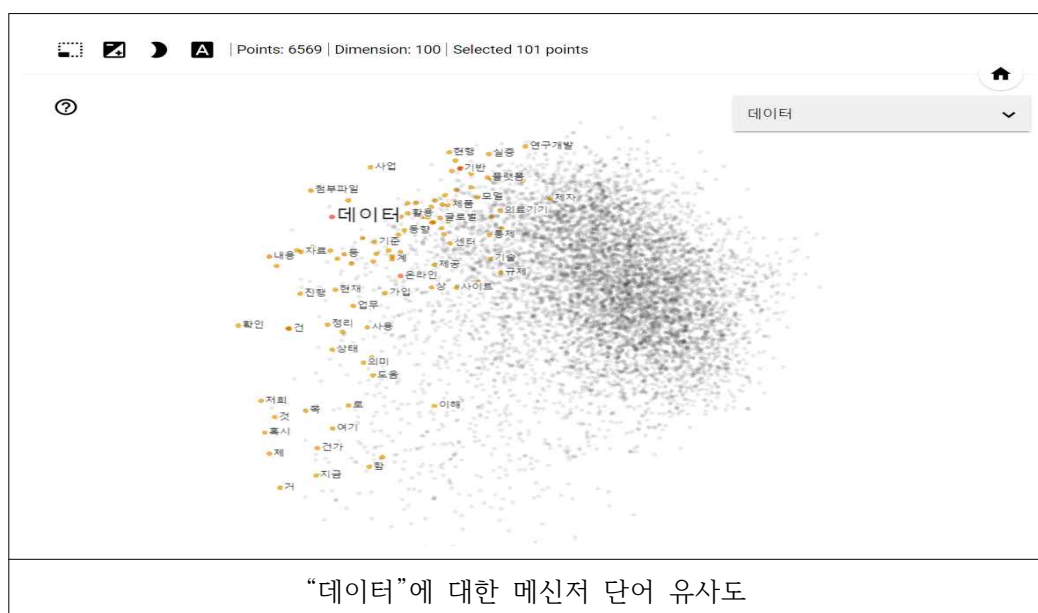
- 정보관리담당관과 메신저를 많이 주고받는 부서를 기준으로 다시 분석을 해보았다.
- 부서별로 쪼개서 보게 되면 특징을 볼 수 있다.
- 가장 눈에 띄는 IT 지원센터(행정정보)의 경우 “메일”, “문의”, “요청” 이
런 단어가 주로 보인다.
- 부서별로 쪼개서 보아도 “최영애”라는 단어가 많이 보이고 있다. 최영애
주무관님이 많은 부서와 메시지를 주고받는 것을 알 수 있다.

□ 단어 유사도

- 코사인 유사도¹⁾로 해당단어와 어떤 단어와 유사한지, 해당 단어와 어떤 단어가 같이 쓰이는지 알 수 있다.
- 예를 들어 “데이터” 라는 단어를 입력했을 때 메신저에서 “데이터”와 가장 유사한 단어를 출력해준다.



- “기반”, “온라인”, “빅데이터”, “플랫폼” 이런 단어가 주를 이루고 있다. 메신저를 사용하면서 “데이터 기반”, “데이터 플랫폼” 이런 단어들을 많이 쓰고 있는 듯하다.



1) 단어(벡터화)를 2차원 공간상에 올려두고 코사인값을 활용하여 단어간의 유사한 정도를 계산

□ 분석 결과 및 추가사항

○ 분석 결과 및 추가사항

- 부서수가 많아 모든 부서를 다 분석 할 수 없었다. 정보관리담당관 중심으로 분석을 해본 결과 IT관련 부서들과 많은 소통이 있다고 판단되고 소통과정에서 최영애 주무관님이 많이 언급됐다.
- 메신저를 주고받은 부서별로 업무적 특성이 드러났다. 추후에 부서별로 많이 나타난 단어를 모아서 불편사항이나 업무에 필요한 것들을 미리 파악이 될 수 있겠다.
- 일주일치 데이터밖에 활용을 못했지만 조금 더 많은 데이터가 있다면 주, 월 별로 해당 부서의 이슈 및 메신저 이용 패턴 등을 확인 할 수 있겠다.
- 다음 분석에는 업무개선을 중점적으로 파악하고 해결하기 위해서 메신저 분석 시행하고 실제 메신저 분석을 바탕으로 어떠한 문제를 해결하는 단계까지 접근하겠다.