

멸종위기 야생생물 민원 텍스트 마이닝 연구
- LDA 토픽 모델링과 네트워크 분석을 통한 주요 이슈 발굴 -

김나영¹⁾²⁾ · 남희정¹⁾³⁾ · 박용수⁴⁾

¹⁾국립생태원 멸종위기종복원센터 전문위원 · ²⁾서울대학교 대학원 협동과정 조경학 학생 ·
³⁾국민대학교 대학원 산림자원학 학생 · ⁴⁾국립생태원 멸종위기종복원센터 팀장

A Text Mining Study on Endangered Wildlife Complaints
- Discovery of Key Issues through LDA Topic Modeling and Network Analysis -

Kim, Na-Yeong¹⁾²⁾ · Nam, Hee-Jung¹⁾³⁾ and Park, Yong-Su⁴⁾

¹⁾National Institute of Ecology, Research Center for Endangered Species,
²⁾Interdisciplinary Program in Landscape Architecture, Seoul National University,
³⁾Dept. of Forest Resources, Graduate School, Kookmin University.

ABSTRACT

This study aimed to analyze the needs and interests of the public on endangered wildlife using complaint big data. We collected 1,203 complaints and their corresponding text data on endangered wildlife, pre-processed them, and constructed a document-term matrix for 1,739 text data. We performed LDA (Latent Dirichlet Allocation) topic modeling and network analysis. The results revealed that the complaints on endangered wildlife peaked in June-August, and the interest shifted from insects to various endangered wildlife in the living area, such as mammals, birds, and amphibians. In addition, the complaints on endangered wildlife could be categorized into 8 topics and 5 clusters, such as discovery report, habitat protection and response request, information inquiry, investigation and action request, and consultation request. The co-occurrence network analysis for each topic showed that the keywords reflecting the call center reporting procedure, such as photo, send, and take, had high centrality in common, and other keywords such as dung beetle, know, absence and think played an important role in the network. Through this analysis, we identified the main keywords and their relationships within each topic and derived the main issues for each topic. This study confirmed the increasing and

First author : Kim, Na-Yeong, National Institute of Ecology, Research Center for Endangered Species,
Tel : +82-54-680-7256 , **E-mail** : nynayeong@gmail.com

Corresponding author : Park, Young-Su, National Institute of Ecology, Research Center for Endangered Species,
Tel : +82-54-680-7250, **E-mail** : muskdeer@nie.re.kr

Received : 26 October, 2023. **Revised** : 20 December, 2023. **Accepted** : 15 November, 2023

diversifying public interest and complaints on endangered wildlife and highlighted the need for professional response. We also suggested developing and extending participatory conservation plans that align with the public's preferences and demands. This study demonstrated the feasibility of using complaint big data on endangered wildlife and its implications for policy decision-making and public promotion on endangered wildlife.

Key Words: *Endangered wildlife, Endangered wildlife policy, Civil data topic modelling, Network analysis, Complaint data about endangered wildlife*

I. 서론

생물다양성은 생태계의 건강과 기능을 유지하기 위한 필수 요소이다. 그러나 인간 활동으로 인해 생물다양성이 지속적으로 감소하고 있다. 이는 생태계의 회복력과 지속가능성을 저하시키는 위협 요인으로 국제사회는 이러한 위협을 막기 위해 공동 목표와 행동 계획을 수립하고 있다 (Secretariat of the Convention on Biological Diversity, 2020). 2022년 12월에는 쿤밍-몬트리올 글로벌 생물다양성 프레임워크가 채택되어 2030년까지 생물다양성의 감소를 역전시키고, 2050년까지 인간과 자연의 조화를 달성하기 위한 목표와 계획을 제시하였다(Ministry of Environment, 2022; Lim et al., 2023). 우리나라에서도 제5차 국가생물다양성 전략(2024-2028)을 수립하여 멸종위기 야생생물의 보호와 복원을 포함한 다양한 노력을 하고 있다(Ministry of Environment, 2023). 전략의 효과적인 실행과 목표 달성을 위해, 멸종위기 야생생물의 보전과 복원에 대한 국민의 관심과 요구를 반영하는 정책 결정과 이행이 필요하다.

다양한 정책 분야에서 행정의 객관성과 효율성 증진을 위해 빅데이터를 활용한 증거 기반 정책수립이 진행되는 가운데(Duberry, 2019) 민원 빅데이터를 활용한 텍스트 마이닝 분석은 시민의 수요와 실제 공공서비스에 대한 경험 등에 초점을 맞추어 전반적인 사회 현상을 구체적으로 살펴보고 이를 통한 적시 정책 수립을 가능하게 한다(Chung et al., 2016; Ryu et al., 2018; Park & Lee, 2020;

Jiao et al., 2021). 멸종위기 야생생물 관련 정책 제안 및 운영에 활용가능한 기초 연구로 멸종위기 야생생물에 대한 대국민 인식조사(National Institute of Ecology, 2020; 2022)와 연구 동향 분석(Do et al., 2020; Yoon et al., 2021; Chae et al., 2022), 민원 빅데이터 분석을 통한 환경 이슈, 도시 공원 문제점 도출(Choi, 2016; Jin et al., 2019; Sung & Lee, 2022), 국내외 지역생물다양성 전략 비교를 통한 정책 방향성 제안 연구(Lee & Sung, 2018; Choi et al., 2022)가 실시된 바 있으나 멸종위기 야생생물 관련 내용은 한정적이었고 이를 통해 정책을 제안과 적용에 어려움이 있었다. 또한 멸종위기 야생생물 관련 선행 연구의 경우, 연구의 방향이 특정종 혹은 분류군의 생태와 서식 환경, 서식지 모델, 복원, 대체서식지, 사육 혹은 증식 등의 일부 주제에 집중되어 있어 실제 멸종위기 야생생물 관련 국민의 관심사를 살펴보기나 반영한 연구는 미비하였다. 한편, 정부 중심의 민원 빅데이터 분석 플랫폼 「범정부데이터 분석시스템」, 「한눈에 보는 민원 빅데이터」, 언론사 뉴스 빅데이터 분석 플랫폼 「빅카인즈」 등 빅데이터를 활용한 정책 제도 개선 추진과 지원 또한 활발하게 이루어지고 있다. 이에 실제 국민의 관심과 목소리를 반영한 멸종위기 야생생물 보전 관련 제도 개선 및 정책 마련에 뒷받침이 될 수 있는 실증 연구, 분석이 요구된다.

본 연구는 멸종위기 야생생물 통합콜센터를 통해 수집된 민원 텍스트 데이터를 활용하여 국민의 멸종위기 야생생물 관련 주요 관심사를 도

Table 1. Text mining analysis data details

Distinction	Description	Number of data(Number of case)				
		Text message	Telephone	E-mail	i-net homepage board	SNS message
Collected text data	Textualized data that has been processed through basic cleanup	782	791	272	191	96
		2,132				
Complaint text data	After pre-processing, data ready for text mining analysis	588	742	162	163	84
		1,739				

출하고, 그 변화상과 특징을 파악하고자 한다. 전체 접수 민원 사례에 대한 탐색적 분석을 통해 민원의 추이를 분석하고, 텍스트 마이닝 기법 중 하나인 토픽모델링 기법-잠재 디리클레 할당(Latent Dirichlet Allocation; 이하 LDA)을 활용하여 각 민원 텍스트 데이터를 질차적 확률 분포에 따라 토픽 분류하였다. 마지막으로 분류된 토픽별 네트워크를 분석하여 주요 쟁점을 제시하였다. 이러한 민원 빅데이터 분석 결과를 통해 국민이 공감할 수 있는 멸종위기 야생생물 관련 보호, 정책 개발 및 서비스 제공을 위한 방향성 제시에 기여하고자 한다.

II. 연구방법

1. 연구 데이터의 수집과 범위

멸종위기 야생생물 통합콜센터는 멸종위기 야생생물 신규 서식지 발굴과 보전 활동 및 정확한 정보 제공을 목적으로 개설되었다. 2020년 4월 정식 운영 이후 국민의 멸종위기 야생생물 보전 활동의 소통 창구로 멸종위기 야생생물 전국 분포 조사·신규 서식지 발굴, 멸종위기 야생생물 서식지 특성 연구, 멸종위기 야생생물 목록 개정 등 다양한 분야에서 민원 데이터가 활용된 바 있다. 2023년 9월 운영 체계를 개편하여 멸종위기 야생생물 발견 제보 창구로 활용하고 있다. 본 연구는 2020년 5월 1일부터 2023년 5월 31일까지 접수된 총 1,203건의 민원을 대상으로 하였다. 『민원 처리에 관한 법률』에서 정의하는 일반민원(법정민원, 질의민

원, 건의민원, 기타민원), 복합민원으로 한정하여 데이터를 수집하였다(CIVIL PETITIONS TREATMENT ACT, 2022). 복수의 민원 접수창구를 통해 수집된 텍스트 데이터를 민원 목적과 접수 내용에 따라 2,132건의 문서로 정리하였으며 전처리 작업 수행 후, 최종적으로 1,739건의 문서를 텍스트 마이닝 분석 연구에 활용하였다. 민원 사례 중 전화로 수집된 음성 데이터는 네이버 클로바 노트의 음성 인식-텍스트 변환 기능을 통해 텍스트 변환 후 음성과 대조, 수정 작업을 거쳐 텍스트 데이터 형태로 가공하였다. 휴대전화 문자 메시지, 모바일 메신저(카카오톡), 전자 메일 및 포털 홈페이지 게시판을 통해 수집된 민원 또한 기본 정제 과정을 거쳐 텍스트 데이터 형태로 가공하였다. 동일 제안자가 복수의 매체를 통해 같은 내용의 민원을 제기한 경우, 시간 순서로 중복 건을 제거하였으며 세부 내용이 다르거나 추가 정보가 주어지는 경우, 복수의 데이터를 모두 활용하였다. 민원 텍스트 데이터는 실제 민원인의 관심과 의도 등을 파악하기 위해 민원 사례별 분류하지 않고 수집된 민원 텍스트 정리 데이터(문서)를 정제, 분석하였으며 답변 내용은 포함하지 않았다. 또한, 해석 가능성을 고려하여 전처리 후 키워드가 3개 이상인 텍스트를 대상으로 분석을 수행하였다. 활용 데이터에 대한 상세 내용은 [Table 1]과 같다.

2. 텍스트데이터 전처리와 주요 연구 방법

1) 텍스트 마이닝 분석을 위한 전처리

텍스트 데이터는 R 기반의 KoNLP 자연어 처

리 패키지를 활용하여 전처리 작업을 수행하였다. 민원 사례 정보 수집 시 민원인의 이름, 연락처와 같은 개인정보는 사전 블라인드 처리되었으며, 영문은 한글로 변환하고, 숫자, 오타 등 분석 과정에서 불필요한 요소들은 수정하거나 삭제하는 등 정제 과정을 거쳤다. 또한 ‘안녕하세요’, ‘감사합니다’, ‘수고하세요’ 같은 인사말이나 ‘이거’, ‘저거’, ‘여기’ 등과 같은 지시어는 불용어로 설정하여 분석 시 제외하였고 추가로, 멸종위기 야생생물의 특수성을 고려하여, 지역 관련 키워드들도 제거하였다. 유의어는 출현수가 많은 하나의 키워드로 통일하여 같은 의미인 키워드가 복수로 나타나지 않게 하였으며 최소 단어수를 설정하여 분석이 더욱 정확하게 이루어질 수 있도록 하였다(Park, 2020). 최소 단어수는 2로 설정하여 ‘삼’, ‘매’, ‘산’, ‘강’처럼 2자 이하의 단어는 영문으로 변경하거나 2자 이상의 다른 단어로 대체하여 처리하였다. 추가로, 도출된 데이터 인덱스를 확인하여 전체 문서 내 3회 미만 등장한 키워드의 경우 삭제하여 분석과 분류의 편의성을 높이고자 하였다. 빈도수 분석과 토픽 클러스터링 과정에서는 명사를 추출하여 전처리하고 분석하였으며 분류된 토픽의 네트워크 분석 과정에서는 명사, 형용사, 동사를 추출하여 전처리하고 분석하였다.

2) 탐색적 데이터 분석과 TF-IDF 분석

연구 대상으로 선정된 민원 1,203건에 대해 제보수 변화, 주요 관심 대상종 비율 등 기초 통계를 포함하는 탐색적 데이터 분석을 하였다. 또한 각 데이터 세트의 주요 특징을 요약하여 전반적인 민원 텍스트 데이터에 대한 기술통계를 수행하였다. 이 때, 수집 시기별 변화를 알아보기 위해 동일한 3개의 기간으로 데이터를 나누고 분석에 활용하였다. 구체적으로는 기간 A를 2020년 5월부터 2021년 4월로 설정, 기간 B를 2021년 5월부터 2022년 4월로 설정, 기간 C를 2022년 5월부터 2023년 4월로 지정하였으며

2023년 5월 수집된 데이터의 경우 수집 기간이 달라 기간으로 나누어진 데이터에는 포함되지 않았다. TF-IDF(Term Frequency - Inverse Document Frequency; 이하 TF-IDF) 분석을 수행하였다. TF-IDF 분석은 복수의 문서 내 어떠한 단어의 중요도를 파악하는데 사용되며(Lim, 2014) 단어의 빈도와 역 문서 빈도를 사용하여 단어마다 중요한 정도에 따라 가중치를 부여한다. TF-IDF 분석은 빈도수 분석과 TF-IDF 분석 결과를 통해 주요 키워드를 도출하고 기간, 민원 수집원별 주요 키워드 빈도수 변화를 살펴보는 등 추후 토픽 모델링과 네트워크 분석 시 데이터의 이해를 돕기위한 기초 자료로 활용하였다.

3) LDA 토픽 모델링과 네트워크 분석

본 연구에서는 잠재 디리클레 할당(latent dirichlet allocation; LDA)모형을 활용하여 수집된 데이터의 토픽 클러스터링 분석을 수행하였다. 잠재 디리클레 할당 모형은 하나의 문서가 여러 토픽의 혼합으로 구성되고 문서의 주제에 따라 단어의 분포가 결정된다는 가정 하에 문서 전체의 토픽, 각 문서별 개별 토픽 비율, 각각의 토픽에 대한 단어 분포를 확률적으로 추정, 계산한다(Blei, 2012). 토픽의 수는 복잡도 지표 Griffiths2004, CaoJuan2009, Deveaud2014와 모형 검진 계수 의미 일관성(semantic coherence), 지속성(heldout likelihood), 잔차(residual) 등을 고려하여 범위를 지정하고 주제 모형 적합을 통해 설정하였다. 범위 내 모델 개수에 대해 개별 기계학습을 수행하여 도출 토픽 결과를 키워드-대표 문서의 일치성 등을 종합해 해석 가능성이 가장 큰 모델을 최종적으로 선택하였다(Wallach et al., 2009; Yang et al., 2021; Lee, 2021). 문서별 토픽을 할당한 이후, 동시출현 빈도를 이용하여 단어의 관계를 네트워크 형태로 표현하는 동시 출현 네트워크 분석을 통해 토픽별 네트워크를 도출하고 네트워크에서 중심 단어와 주제 군집을 찾을 수 있도록 하였다. 이를 통해 노드 즉, 키워드 간의 관계로 형성되어 있는 구조적

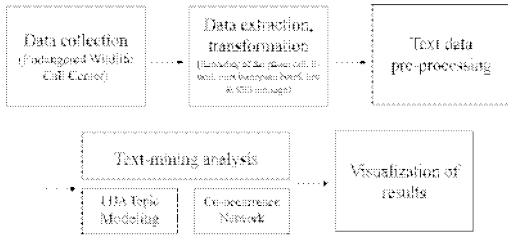


Figure 1. Process of research

성격을 정량적 이해할 수 있고 네트워크를 통해서 구조적 변수에 대한 행위를 설명할 수 있다.

III. 결과 및 고찰

1 멸종위기 야생생물 통합콜센터 민원 데이터 분석

2020년부터 5월부터 2023년 5월까지의 민원은 총 1,203건으로 연 평균 300.75건으로 계산되며 2020년 218건, 2021년 329건, 2022년 455건, 2023년 201건으로 확인된다. 동일 기간 대비 민

원 발생량을 비교해보면, 281건(2020.5.~2021.4.; 이하 기간 A), 354건(2021.5.~2022.4.; 이하 기간 B), 436건(2022.5.~2023.4.; 이하 기간 C)으로 증가하였다. 민원 발생 수를 월별로 비교한 [Figure 2]를 보면, 연도별 소폭 차이가 있으나 여름철인 6~8월 민원 접수가 확연하게 증가하고 겨울에는 민원이 감소하는 양상을 보여 계절적 편향이 존재하는 것을 확인할 수 있다. 야외 활동이 많은 시기에 민원이 집중되는 경향은 야생생물의 발견, 제보 확률이 상대적으로 높아져 나타나는 현상으로 해석할 수 있다. 수집 매체별 민원 발생량을 정리한 [Table 2]를 살펴보면 전화를 통한 문의가 가장 큰 비율을 차지하며 민원 데이터 수집 채널 확장이 있었던 2021년(포털 홈페이지), 2022년(모바일 메신저)에 전체 민원 제보 수가 소폭 증가한 것으로 보인다. [Figure 2]에서 2023년 4월~5월 급격한 제보 증가를 확인

Table 2. Changes in the number of complaints received by media

Year / Media	E-mail	Text message	Telephone	SNS message	I-net homepage board
2020	21	-	197	-	-
2021	39	102	151	-	37
2022	31	111	189	24	100
2023	24	57	84	16	20

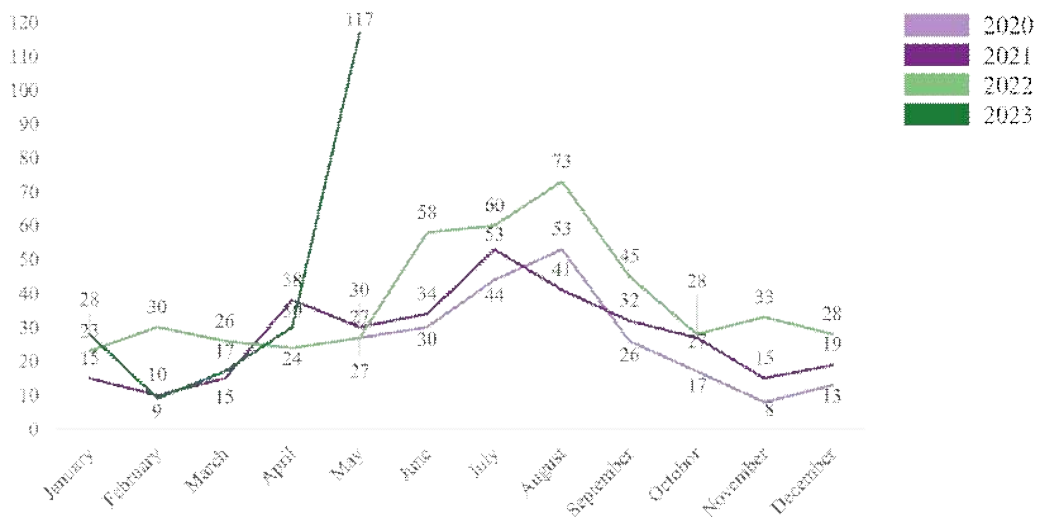
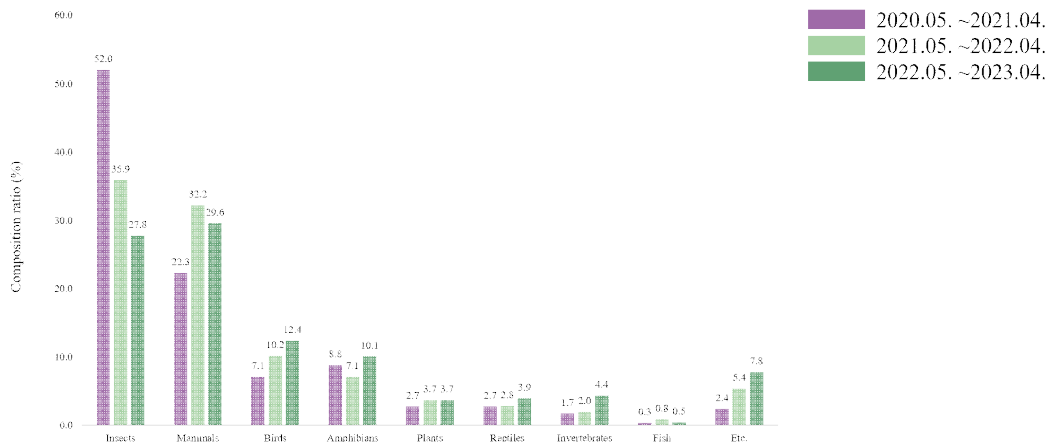


Figure 2. Monthly changes in the number of complaints

Table 3. Changes in the number of complaint target taxonomic group by year

Taxonomic group / Year	2020	2021	2022	2023
Amphibians	21	25	46	9
Birds	15	25	58	15
Fish	-	3	2	1
Insects	141	134	110	99
Invertebrates	3	7	16	9
Mammals	22	97	152	54
Plants	5	14	16	2
Reptiles	7	9	18	6
Etc.	4	15	37	6

**Figure 3.** Changes in the composition target taxonomic group ratio of complaints by period

할 수 있는데 이 시기에 홈페이지 발견 제보 게시판 통합으로 민원이 일시에 접수되었고 단순 발견 제보의 후속 민원이 다수 접수된 결과라 할 수 있다.

전체 접수된 민원 중 곤충과 포유류에 대한 민원이 67.25%를 차지하며 연도별 주요 관심 분류군 비율의 변화가 나타났다[Table 3]. 동일 기간 대비 분류군별 민원 발생량을 비교한 [Figure 3]을 보면, 기간 A에는 곤충에 집중되어 있던 민원 비율이 시간이 지나면서 포유류, 조류, 양서류 등에 분산되는 것을 알 수 있다. 특히, 조류, 파충류, 무척추동물에 대한 민원 데이터는 소폭이지만 지속해서 증가하는 추세를 보였다. 총 247종에 대한 민원이 수집되었으며 그중 187종

(62.5%)이 멸종위기 야생생물이 아닌 종으로 단순 정보문의나 지역 절멸로 보고된 호랑이, 표범, 소똥구리 등에 대한 발견 제보 중 실체 확인이 불가능한 경우도 포함되었다. 이는 멸종위기 야생생물에 대한 민원을 해결하기 위한 창구로써 멸종위기 야생생물 통합콜센터를 운영하였으나 일반종에 대한 국민들의 관심과 보호에 대한 민원이 많았음을 보여준다. 특정 종에 대한 민원 중 소똥구리에 대한 발견 제보, 정보문의 등 민원이 365건, 30.34%로 계산되어 전체 민원 중 가장 큰 부분을 차지하였다. 연도별 118건, 93건, 75건, 82건으로 2021년 가장 큰 감소세를 보인 후, 소폭 증가 추세를 보였다.

질의 사항으로는 크게 국제적멸종위기야생생

Table 4. Frequency of keywords appearing in the complaint data

keyword	count	keyword	count
photo (사진)	516	inquiry (문의)	103
dung beetle (소똥구리)	355	protect (보호)	101
verification (확인)	299	environment (환경)	99
phone call (전화)	269	habitat (서식지)	96
object (개체)	211	Ministry of Environment (환경부)	92
contact (연락)	145	narrow-mouth frog (맹꽁이)	91
text message (문자)	132	area (지역)	89
otter (수달)	130	possible (가능)	87
person (사람)	127	accurate (정확)	84
location (위치)	125	inhabit (서식)	81
curiosity (궁금)	118	if (만약)	79
yellow-throated marten (담비)	118	vicinage (근처)	77
animal (동물)	113	ecology (생태)	76
place (장소)	112	thought (생각)	75
necessary (필요)	104	video (영상)	75

물 수·출입, 생태자연도 등급 조정 같은 정책 관련 내용이 많았으며 관련 지자체와 유역·지방환경청 관계자 혹은 일반인이 멸종위기종 발견 이후 절차에 관해 문의하는 사례가 연도별 소폭 증가하는 경향을 보이는 것으로 나타났다. 추가로 이러한 기타문의에 대한 민원 주체는 지자체와 유역·지방환경청 관계자가 일반 민원인보다 많은 것으로 확인되었다.

2. 민원 텍스트 데이터 빈도수 분석

텍스트 전처리 후, 주요 키워드 탐색을 위해 1,739개의 텍스트 데이터에 대하여 명사만을 추출하여 빈도수 분석을 수행하였다. 민원 전체 내용에 대해서 자주 언급되는 키워드의 출현 빈도수 확인해 본 결과, [Table 4]와 같았다. 전체 6,278개의 키워드 중 출현 빈도수 상위 10%에 해당하는 키워드의 빈도수 합이 총빈도수의 61.19%를 기록해 수집된 데이터가 주요 키워드를 중심으로 구성되어 있다는 것을 추측할 수 있다. 추가로 출현 빈도수 상위 50개의 빈도수 합은 총빈도수의 22.9%를 차지하는 것으로 나타났

다. 멸종위기 야생생물 통합콜센터의 특성상 종 동정 및 서식지 확인을 위한 ‘사진’, ‘전화’, ‘문자’, ‘동영상’, ‘메일’과 같은 현장의 상황 전달 연관 키워드가 언급량이 크며 핵심어로 작용하는 것을 알 수 있다. 멸종위기 야생생물 특정 종 관련 키워드로는 ‘소똥구리’, ‘수달’, ‘담비’, ‘맹꽁이’가 빈도수 상위 키워드로 나타나 4종에 대한 민원 관심 정도가 상대적으로 크다고 볼 수 있으며 그 외 ‘애기뿔소똥구리’, ‘샬’, ‘금개구리’, ‘남생이’, ‘여우’의 빈도수가 각각 41, 40, 22, 21, 21회였다. 또한, ‘확인’, ‘필요’, ‘보호’, ‘환경’, ‘서식지’, ‘서식’, ‘생태’가 빈도수 상위 키워드로 포함되었고 ‘환경부’, ‘유역환경청’, ‘기관’, ‘지자체’ 키워드와 더불어 현재 상황에 대한 대처·대응 요구 관련 키워드인 ‘신고’, ‘조사’, ‘복원’, ‘도움’, ‘조치’, ‘보존’, ‘문제’, ‘처리’, ‘포획’, ‘요청’, ‘민원’, ‘허가’, ‘진행’이 평균 출현 빈도수 45.85회로 계산되었다.

민원 수집 매체별 빈도수 상위 50위 내에서 ‘사진’, ‘확인’, ‘수달’, ‘소똥구리’, ‘담비’, ‘궁금’, ‘개체’, ‘위치’, ‘영상’, ‘맹꽁이’, ‘동물’, ‘문의’,

Table 5. Keyword TF-IDF analysis results by period

Period A (2020.05. ~ 2021.04.)				Period B (2021.05. ~ 2022.04.)				Period C (2022.05. ~ 2023.04.)			
keyword	n	tf	tf_idf	keyword	n	tf	tf_idf	keyword	n	tf	tf_idf
cultivation (경작)	3	1.00	8.74	question (질문)	3	0.13	0.96	weekend (주말)	3	0.23	1.70
manchurian weasel (무산쇠족제비)	3	1.00	8.74	beginning (시작)	3	0.13	0.96	expansion (확대)	3	0.23	1.70
chinese weatherfish (미꾸라지)	3	1.00	8.74	plan (계획)	3	0.14	1.09	high school (고등학교)	3	0.25	1.84
change (변화)	3	1.00	8.74	dog (개)	3	0.17	1.23	give (드리다)	3	0.25	1.91
site inspection (실사)	3	1.00	8.74	call (울음소리)	3	0.18	1.35	bioresource (생물자원)	3	0.25	1.91
information sign (안내판)	3	1.00	8.74	classification (분류)	3	0.19	1.43	sunday (일요일)	3	0.27	2.01
now (이제)	3	1.00	8.74	scheduled (예정)	3	0.19	1.43	study (공부)	3	0.27	2.09
petition (청원)	3	1.00	8.74	Inermis (고라니)	3	0.20	1.47	remember (기억)	3	0.30	2.21
grass (풀)	3	1.00	8.74	countermeasure (대책)	3	0.20	1.53	nationwide (전국)	3	0.30	2.21
news (뉴스)	4	1.00	8.74	road kill (로드킬)	3	0.20	1.53	send (전송)	3	0.30	2.21
carp (잉어)	4	1.00	8.74	use (사용)	3	0.20	1.53	mountain (산)	3	0.30	2.29
bean goose (큰기러기)	4	1.00	8.74	region (지방)	3	0.20	1.53	live (생활)	3	0.33	2.45
environmental organization (환경단체)	5	1.00	8.74	kori salamander (고리도롱뇽)	3	0.21	1.64	similar (유사)	3	0.33	2.45
know (알다)	7	0.88	7.05	discoverer (발견자)	3	0.25	1.84	agrochemical (농약)	3	0.33	2.55
stuffy (답답)	5	0.83	6.71	reward (보상)	3	0.25	1.84	behind (뒤)	3	0.33	2.55
front (앞)	4	0.80	6.44	department (부서)	3	0.23	1.86	neglect (방치)	3	0.33	2.55
past (과거)	3	0.75	6.04	rare (희귀)	3	0.23	1.86	despite (불구)	3	0.33	2.55
farmland (농경지)	3	0.75	6.04	health (건강)	3	0.25	1.91	part (일부)	3	0.33	2.55
right now (당장)	3	0.75	6.04	continuously (계속)	3	0.25	1.91	flower garden (화단)	3	0.38	2.76
nest (둥지)	3	0.75	6.04	nation (국민)	3	0.25	1.91	space (공간)	3	0.38	2.87
drone (드론)	3	0.75	6.04	registration (등록)	3	0.25	1.91	military camp (군부대)	3	0.38	2.87
seperate (별개)	3	0.75	6.04	staff (직원)	3	0.25	1.91	snail (달팽이)	3	0.38	2.87
compensation (보상금)	3	0.75	6.04	sunday (일요일)	3	0.27	2.01	classification (분류)	3	0.38	2.87
fence (울타리)	3	0.75	6.04	summer (여름)	3	0.25	2.01	excrements (분변)	3	0.38	2.87
scrutiny (주시)	3	0.75	6.04	official document (공문)	3	0.27	2.09	amur region (아무르 지역)	3	0.38	2.87
environmental preservation (환경 보전)	3	0.75	6.04	education (교육)	3	0.27	2.09	propagation (증식)	3	0.38	2.87
county office (군청)	4	0.67	5.37	farm (농장)	3	0.27	2.09	portal (포털 사이트)	3	0.38	2.87
resources (자원)	4	0.67	5.37	yellow-throated marten (담비)	3	0.27	2.09	leopard (표범)	3	0.38	2.87
fresh water (담수)	3	0.60	4.83	rock (바위)	3	0.27	2.09	appearance (생김새)	3	0.38	3.02
actual situation (실태)	3	0.60	4.83	direction (방향)	3	0.27	2.09	unclear (불명확)	3	0.43	3.15

‘서식’, ‘보호’, ‘주변’ 키워드가 공통으로 확인되었다. 다만, 분석 시 모바일 메신저, 포털 홈페이지 게시판은 빈도 10회 이상 키워드가 각각 5개, 10개로 계산되어 출현 빈도 10회 미만 키워드 비율이 높은 것으로 확인되었다. 가장 많은 민원 데이터를 수집한 전화와 문자에서 33회 이상 공통으로 언급된 키워드는 ‘소똥구리’, ‘사진’, ‘확인’, ‘개체’, ‘위치’로 나타났다. 수집 연도별 상위빈도 키워드 50개의 변화를 살펴보면, 주요 키워드가 공통으로 출현하는 것을 알 수 있었다. 이러한 수집 연도별 공통 상위빈도 키워드들은 연관 키워드들과 함께 의미 해석이 가능하며 연도별 순위 변화를 확인할 수 있다. 연관 키워드 그룹은 ① 민원-제보 시 요청되는 정보와 연관되는 ‘사진’, ‘전화’, ‘문자’, ‘위치’, ‘장소’, ‘지역’,

‘시간’, ② 민원의 목적과 연관되는 ‘확인’, ‘연락’, ‘문의’, ‘궁금’, ‘정확’, ‘가능’, ‘필요’, ‘환경부’, ③ 민원의 대상이 되는 ‘생물’, ‘동물’, ‘소똥구리’, ‘담비’, ‘맹꽁이’, ‘개체’, ④ 상황에 대한 설명으로 주로 쓰이는 ‘환경’, ‘분변’, ‘사람’, ‘근처’, ‘서식지’ 4개의 그룹으로 주요 키워드별 출현 빈도수 변화를 확인할 수 있다. 이 외 ‘판매’, ‘판단’, ‘기사’ 키워드가 2023년 신규 상위빈도 키워드로 확인되었으며 ‘조사’, ‘보호’, ‘주변’ 키워드는 2022년까지 빈도수 상위 50위 내에서 연속적으로 확인되었으나 2023년 빈도수 상위 50위 내에 포함되지 않았다.

기간 B, C에 공통적으로 ‘일요일’이 중요 키워드로 나타났으며 기간 C에서 중요도 순위가 상승하는 것을 알 수 있다. 기간별 중요도가 높

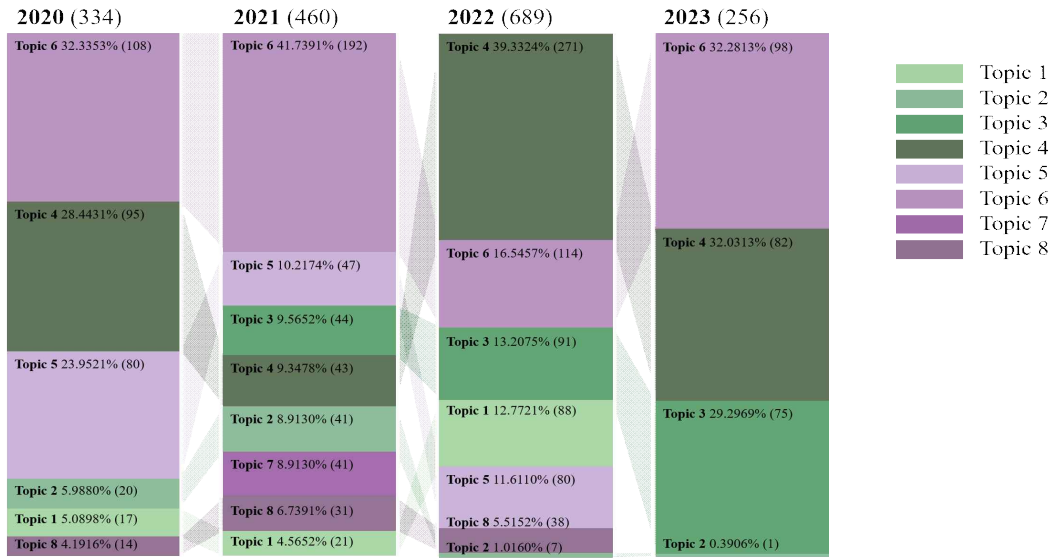


Figure 6. Variation in the number of complaints by topic according to LDA topic model classification

Table 6. Keywords by topic and expected proportion of total data

Topic	Top 10 terms	Expected topic proportion (γ)
Topic 1	inquiry, life, verification, phone call, photo, wild, necessary, contact, Ministry of Environment, person	0.0978
Topic 2	verification, contact, phone call, investigation, construction, ecology, inquiry, area, data, possible	0.1522
Topic 3	narrow-mouth frog, phone call, sound, curiosity, photo, case, if, verification, protect, environment	0.0878
Topic 4	photo, otter, inhabit, phone call, observation, time, protect, verification, person, environment	0.1251
Topic 5	yellow-throated marten, photo, video, animal, stroll, phone call, otter, help, location, observation	0.1507
Topic 6	photo, dung beetle, phone call, person, text message, verification, contact information, fece, internet, contact	0.2366
Topic 7	location, area, conservation, management, development, problem, field, inhabit, hands-on worker, environment	0.0709
Topic 8	animal, protect, possible, case, photo, national red list, thought, artificiality, help, habitat	0.788

은 키워드 중 특정 종 관련 키워드로는 ‘무산쇠족제비’, ‘큰기러기’, ‘고리도롱뇽’, ‘담비’, ‘표범’, ‘거제외줄달팽이’, ‘울릉도달팽이’, ‘참달팽이’가 확인되었다. 해당 키워드들의 빈도수는 ‘소똥구리’, ‘수달’ 등 주로 언급되는 종의 출현

빈도와 비교해 많지는 않지만 특정종에 대한 민원으로 분류되는 데이터에서 자주 발견되는 것으로 해석할 수 있다.

주요 관심 대상으로 언급된 소똥구리의 경우, 2020년 1순위(110회) 이후 2021년 2순위(107

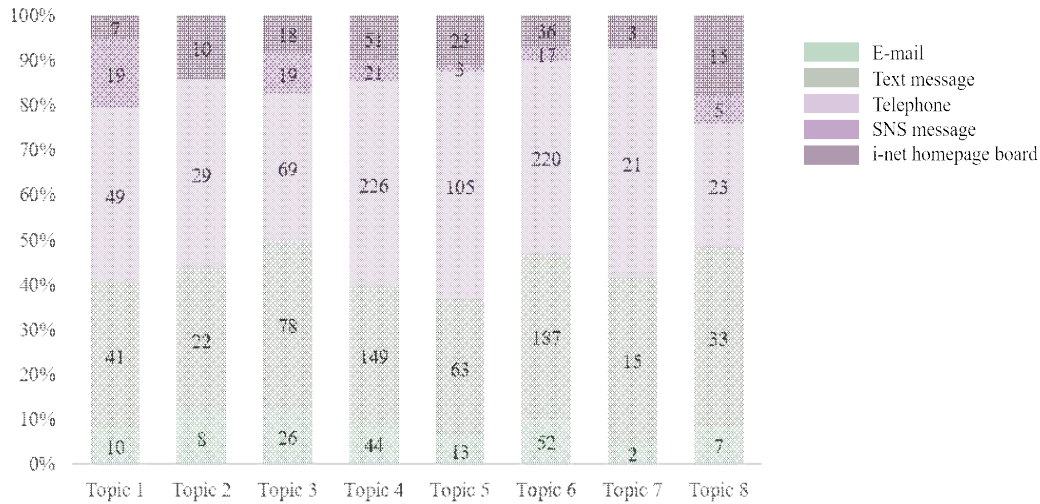


Figure 7. Variation in the number of complaints by topic according to LDA topic model classification

회), 2022년 5순위(83회), 2023년 2순위(55회)로 순위가 변화하였으며 민원 수집 기간과 TF-IDF 값을 고려하였을 때는 기간 A 기간 2.47(220회), B 기간 2.04(99회), C 기간 1.94(94회)로 중요도와 관심도 모두 감소 추세에 있다고 해석할 수 있다. 반면, 2020년 9회 언급, 53순위로 계산된 수달은 2021년 6순위(44회), 2022년 12순위(55회), 2023년 8순위(22회)로 확인되며 민원 수집 기간과 TF-IDF 값을 고려하였을 때는 기간 A 기간 1.75(31회), B 기간 2.60(46회), C 기간 2.89(51회)로 나타나 중요도와 관심도 모두 증가하고 있음을 알 수 있다.

3. LDA 토픽 모델링 분석

전처리 과정을 통해 도출된 32,292개의 텍스트 문체에 나오는 6,278개 키워드와 1,173,509개 토큰을 LDA 토픽 모델링 분석한 결과 8개의 토픽으로 유형화하였다. 토픽별 상위의 단어 확률 분포 값을 가지는 주요 키워드는 [Table 6], [Figure 5] 와 같다. Topic 1, 3, 4, 6은 공통으로 ‘사진’, ‘전화’, ‘확인’ 키워드가 높은 순위에 있는 것을 고려할 때, 민원 상황에 대한 보고와 대응을 요청하는 내용의 민원들이 분류되었을 것

이라 해석할 수 있다. Topic 3, 4는 ‘환경’, ‘보호’ 키워드를 추가로 상위 확률 분포 키워드로 가지는 것을 보아 서식 확인 후 해당 지역 혹은 서식지에 대한 보호를 요구하는 민원들로 분류된 것으로 보인다. 또한 Topic 3, 4, 5, 6의 경우 ‘맹꽁이’, ‘수달’, ‘담비’, ‘소똥구리’가 포함되어 특정 종에 관한 보고와 대응 등에 대한 민원들이 분류된 것으로 판단된다. 해당 종 관련 민원 시 파생되는 키워드들을 함께 찾아볼 수 있다. Topic 2, 7의 경우 ‘공사’, ‘조사’, ‘자료’, ‘확인’, ‘관리’, ‘개발’, ‘문제’와 같은 개발 상황에 따른 키워드들이 상위단어 확률 분포를 보인다. 특히, Topic 3과 Topic 7은, ‘담당자’ 를 상위 확률 분포 키워드로 가지고 있어 다양한 사례별 대응 방안 문의나 즉각 처리 요구 관련 키워드를 가지는 문서가 함께 분류되었을 것으로 판단된다.

전체 민원 데이터(1,739개)에 대해 Topic 6으로 분류되는 민원의 데이터 비율은 29.44%(512건)로 Topic 7로 분류되는 민원의 데이터의 비율이 2.36%(41건)로 가장 낮게 나타났다. 연도별 분류된 토픽 변화를 살펴보면(Figure 6), 전체 민원이 분류되는 토픽 수가 2021년 이후로 감소하는 것을 확인할 수 있다. 수집 매체별로 분류된

토픽 변화를 살펴보면(Figure 7), 문자 메시지가 모든 토픽에서 30% 이상의 비율을 차지한다. 전화를 통해 수집된 민원 데이터는 Topic 4에서 46.03%(226건)로 가장 많이 포함되었고, Topic 5에서 50.72%(105건), Topic 7에서 51.22%(21건)를 차지하여 Topic 4, 5, 7에 큰 영향을 주었다고 할 수 있다. 한편, 모바일 메시지를 통해 수집된 데이터는 Topic 4로 21건이 분류되어 가장 많이 나타났으며, Topic 2, Topic 7로는 분류되지 않았다. E-mail을 통해 수집된 민원 데이터는 Topic 6으로 가장 많은 52건이 분류되었고 포털 홈페이지 게시판으로 수집된 민원은 Topic 4로 가장 많은 수가 분류되었다.

전체 민원 데이터에 대해 Topic 6으로 분류되는 민원의 데이터 비율은 29.44%, 512건이며 Topic 7로 분류되는 민원의 데이터의 비율이 2.36%로 가장 낮게 나타났다. 연도 별 분류된 토픽 변화를 살펴보면, 전체 민원이 분류되는 토픽 수가 2021년 이후로 감소하는 것을 확인할 수 있다. 매체별로 분류된 토픽 변화를 살펴보면, 문자 메시지가 모든 토픽에서 30% 이상의 비율을 차지한다. 전화로 수집된 민원 데이터는 Topic 5에서 105건으로 50.72%, Topic 7에서 21건, 51.22%를 차지하여 해당 Topic에 큰 영향을 주었다고 할 수 있으며, Topic 4로 226건이 분류되어 가장 많은 수를 차지하는 것으로 나타났다. 한편, 모바일 메시지를 통해 수집된 데이터의 경우, Topic 4로 21건이 분류되어 가장 많이 나타났으며, Topic 2, Topic 7로는 분류되지 않았다. E-mail을 통해 수집된 민원 데이터는 Topic 6으로 가장 많은 52건이 분류되었고 포털 홈페이지 게시판으로 수집된 민원은 Topic 4로 가장 많은 수가 분류되었다. 전체 토픽의 변화 양상은 [Figure 6], [Figure 7]에서 확인할 수 있다.

4. 토픽별 네트워크 분석

토픽별 네트워크 분석결과는 [Figure 8], [Figure 9]와 같이 나타났다. Topic 1의 주요 키워

드는 ‘문의’, ‘생물’, ‘확인’, ‘전화’, ‘사진’, ‘야생’, ‘필요’, ‘연락’, ‘환경부’, ‘사람’으로 나타났다. Topic 1의 네트워크 분석 결과 밀도(edge density)는 0.0844, 클러스터링 계수(transitivity)는 0.3267로 계산되었다. 연결중심성이 높은 키워드는 ‘같다’, ‘아니다’, ‘드리다’, ‘들다’, ‘없다’, ‘알다’, ‘환경’, ‘문제’, ‘주변’, ‘보호’ 등이었다. 이를 통해 멸종위기 야생생물 여부 확인(중 동정)으로 설명할 수 있으며 Topic 1으로 분류된 민원 데이터 원본 내용을 살펴보면 멸종위기 야생생물로 의심되는 종에 대한 발견 제보 혹은 확인을 요청하는 민원이 포함되었다. 특히 국민들에게 인지도가 높은 소꿉구리, 삿, 수달, 담비에 대한 중 동정을 요청하는 민원 많았으며 이들의 서식 환경 보호를 위해 민원을 접수하는 경향이 있었다.

Topic 2의 주요 키워드는 ‘확인’, ‘연락’, ‘전화’, ‘조사’, ‘공사’, ‘생태’, ‘문의’, ‘지역’, ‘자료’, ‘가능’으로 나타났다. Topic 2의 네트워크 분석 결과 밀도는 0.1508, 클러스터링 계수는 0.355로 계산되었다. 연결중심성이 높은 키워드는 ‘보내다’, ‘알다’, ‘맞다’, ‘아니다’, ‘드리다’, ‘다르다’, ‘받다’, ‘모르다’, ‘확인’, ‘연락’ 등 이다. 엣지 수가 190으로 전체 네트워크 중 두 번째로 작으나 매개중심성 값이 0.3942로 가장 높은 값을 가지는 것으로 확인되었다. 따라서 Topic 2는 지자체, 경찰서, 소방서 등 기관 협조를 포함한 발견 제보 민원으로 설명할 수 있으며 민원 데이터 원본 내용을 살펴보면 멸종위기 야생생물 발견 제보와 더불어 서식 확인에 대한 공문을 요청하거나 공사 중지 등 강도 높은 보호 활동을 위한 기관 협조 요청 민원도 다수 포함되었다.

Topic 3의 주요 키워드는 ‘맹꽁이’, ‘전화’, ‘소리’, ‘궁금’, ‘사진’, ‘경우’, ‘만약’, ‘확인’, ‘보호’, ‘환경’ 등 순서로 나타났다. Topic 3의 네트워크 분석 결과 가장자리 밀도는 0.065, 클러스터링 계수는 0.2912로 계산되었다. 연결중심성이 높은 키워드는 ‘알다’, ‘같다’, ‘보내다’, ‘많다’, ‘없다’, ‘나오다’, ‘찾다’, ‘여쭙다’, ‘확인’, ‘부탁드

리다, ‘좋다’, ‘방법’, ‘신고’, ‘보호’ 등이다. 따라서 Topic 3는 멸종위기 야생생물 서식지 보호와 혹은 서식지 보호에 대한 불만 민원으로 해석할 수 있었다. 민원 데이터 원본 내용을 살펴보면 음성파일(울음소리)을 포함한 자료로 종 동정을 요청하는 경우가 포함되어 있었으며 특히 맹꽁이로 의심되는 종에 대한 동정 문의와 이후 서식지 보호와 같은 후속 조치 요청 등을 포함한 민원이 많았다.

Topic 4의 주요 키워드는 ‘사진’, ‘수달’, ‘서식’, ‘전화’, ‘관찰’, ‘시간’, ‘보호’, ‘확인’, ‘사람’, ‘환경’ 등 순서로 나타났다. Topic 4의 네트워크 분석 결과 0.049, 클러스터링 계수는 0.3144로 계산되었다. 연결중심성이 높은 키워드는 ‘아니다’, ‘맞다’, ‘모르다’, ‘보내다’, ‘여쭙다’, ‘알다’, ‘같다’, ‘알리다’, ‘환경부’, ‘서식지’, ‘방법’, ‘문

제’, ‘조치’, ‘현장’, ‘지역’ 등으로 나타났다. 다른 토픽 네트워크와 비교하여 ‘서식지’, ‘보호’, ‘환경’, ‘문제’, ‘경우’ 등 다양한 중심성의 크기와 분포를 보이는 세부 키워드들이 확인되었다. 이를 통해 Topic 4는 멸종위기 야생생물에 다양한 정보와 복합질의 민원이 포함된 것으로 보인다. 분류된 민원 데이터 원본 내용을 살펴보면 멸종위기 야생생물로 의심되는 종의 사진을 통한 종 동정 요청, 멸종위기 야생생물(혹은 의심종)에 대한 포획 절차 문의, 멸종위기 야생생물 생태 및 서식 관련 정보 요청, 멸종위기 야생생물 관련 정책 세부 사항에 대한 질의 등이 포함되어 있었다.

Topic 5의 주요 키워드는 ‘담비’, ‘사진’, ‘영상’, ‘동물’, ‘산책’, ‘전화’, ‘수달’, ‘도움’, ‘장소’, ‘목격’ 등 순서로 나타났다. Topic 5의 네트워크

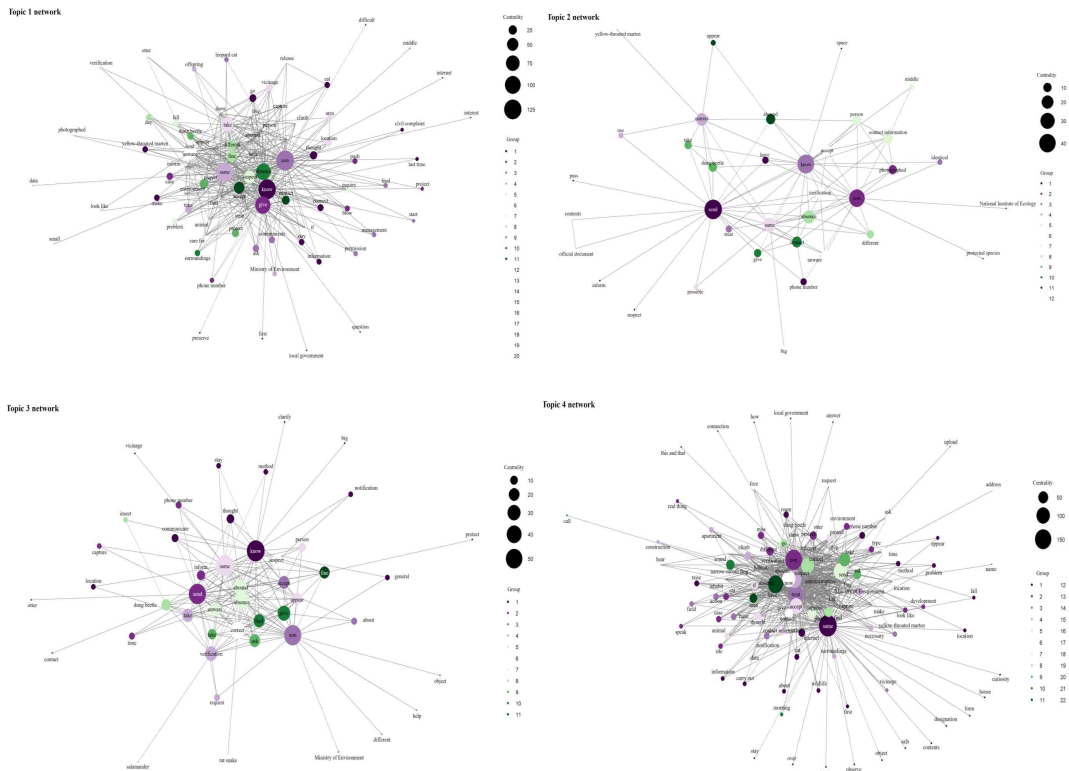


Figure 8. Network based on LDA topic classification (Topic 1 ~ Topic 4)

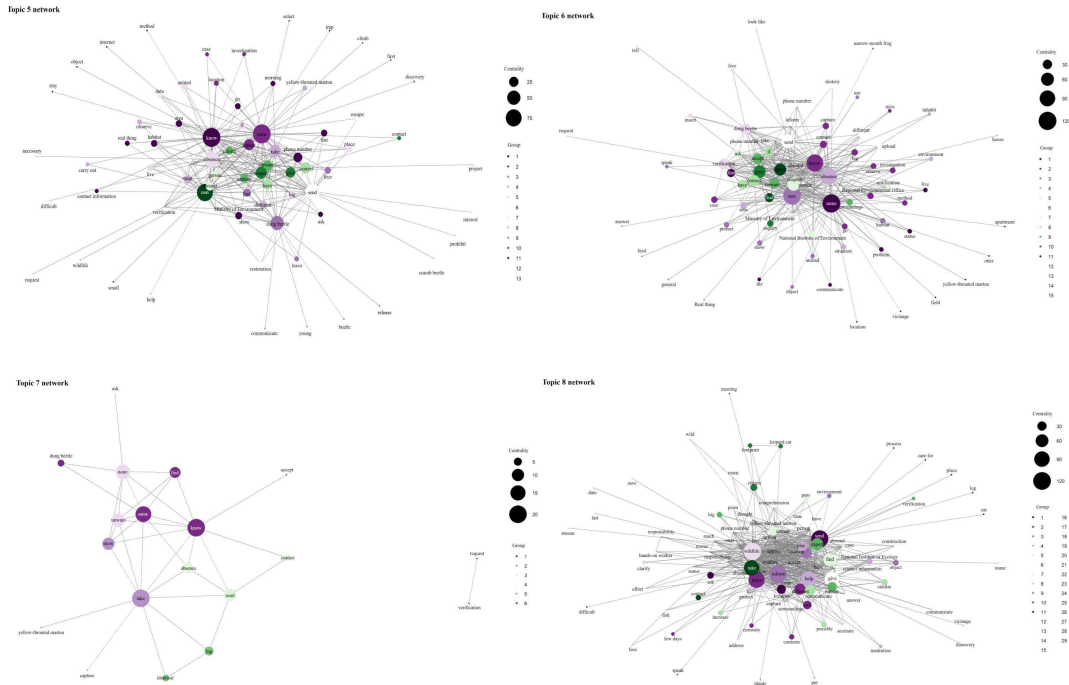


Figure 9. Network based on LDA topic classification (Topic 5 ~ Topic 8)

분석 결과 밀도는 0.0503, 클러스터링 계수는 0.2674로 계산되었다. 연결중심성이 높은 키워드는 ‘알다’, ‘알리다’, ‘없다’, ‘아니다’, ‘나오다’, ‘같다’, ‘모르다’, ‘가지다’, ‘잡다’, ‘보내다’, ‘소똥구리’, ‘분변’, ‘환경부’, ‘확인’, ‘서식지’, ‘조사’ 등으로 나타났다. 해당 토픽 네트워크에는 멸종위기 야생생물인 ‘소똥구리’, ‘애기뿔소똥구리’를 포함, ‘장수풍뎅이’, ‘복원’, ‘탈출’, ‘분변’ 등 곤충 관련 키워드가 확인된다. 이러한 토픽으로 분류된 민원 데이터 원본 내용을 살펴보면 멸종위기 야생생물 여부 확인 요청이 많았으나 과거 멸종위기 야생생물 발견 서식지에 대한 조사 및 정보 제공에 대한 민원도 다수 확인되었다. 멸종위기 야생생물 발견 후 조사, 조치 요구(유역-지방환경청), 과거 발견한 멸종위기 야생생물에 대한 제보 관련 민원으로 설명할 수 있으며

Topic 6의 주요 키워드는 ‘사진’, ‘소똥구리’, ‘전화’, ‘사람’, ‘문자’, ‘확인’, ‘번호’, ‘분변’, ‘전

화번호’, ‘연락’ 등으로 나타났다. Topic 6의 네트워크 분석 결과는 밀도는 0.0409, 클러스터링 계수는 0.2534로 계산되었다. 연결중심성이 높은 키워드는 ‘알다’, ‘드리다’, ‘보내다’, ‘들다’, ‘같다’, ‘없다’, ‘지방환경청’, ‘환경부’, ‘국립생태원’, ‘아파트’, ‘집’, ‘신고’, ‘분변’, ‘포획’, ‘도움’ 등으로 나타났다. 이를 통해 민원인들이 소똥구리 발견에 관심이 많으며 인터넷을 통한 접한 현상금 관련 이슈로 선 포획 후 신고가 이루어지고 있음을 알 수 있다. 추가로 맹꽁이, 담비, 수달에 대한 발견 제보 민원도 함께 분류되었으며 발견 장소가 아파트, 집과 같은 주거지 인근의 생활 범위 내로 한정되는 키워드를 확인할 수 있었다. 이러한 토픽으로 분류된 민원 데이터 원본 내용을 살펴보면 주로 소똥구리 제보가 높은 비율을 차지하여 소똥구리 현상금 문의, 소똥구리와 유사한 곤충 동정 문의 등 발견 제보 등에 대한 민원이 많았다.

Topic 7의 주요 키워드는 ‘위치’, ‘지역’, ‘보존’, ‘관리’, ‘개발’, ‘문제’, ‘현장’, ‘서식’, ‘담당자’, ‘환경’ 등으로 나타났다. Topic 7의 네트워크 분석 결과, 밀도는 0.2164, 클러스터링 계수는 0.2534로 계산되었다. 연결중심성이 높은 키워드는 ‘모르다’, ‘같다’, ‘알다’, ‘없다’, ‘찍다’, ‘보이다’, ‘보내다’, ‘맞다’, ‘아니다’, ‘찾다’ 등으로 나타났다. 엣지 수가 74로 전체 네트워크 중 가장 적으며 단순한 구조를 확인할 수 있다. 봄철 멸종위기 야생생물 발견 제보 민원으로 설명할 수 있으며 이 토픽으로 분류된 민원 데이터 원본 내용을 살펴보면 멸종위기 야생생물 발견, 확인 이후 조치 관련 질문, 멸종위기 야생생물 여부 확인 요청 등에 대한 민원이 많았다.

Topic 8은 시민단체, 지자체 담당자, 유역·지방환경청 등 다양한 이해관계자의 발견 제보 및 자문 요청 민원으로 설명할 수 있으며 주요 키워드는 ‘동물’, ‘보호’, ‘가능’, ‘경우’, ‘사진’, ‘국가적색목록’, ‘생각’, ‘인공’, ‘도움’, ‘서식지’ 등 순서로 나타났다. Topic 8의 네트워크 분석 결과, 밀도는 0.0409, 클러스터링 계수는 0.5로 계산되었다. 연결중심성이 높은 키워드는 ‘아니다’, ‘찍다’, ‘없다’, ‘알다’, ‘좋다’, ‘보내다’, ‘증가하다’, ‘전문가’, ‘도움’, ‘서식지’, ‘전화번호’, ‘담비’, ‘수달’, ‘삶’, ‘발자국’, ‘담당자’, ‘시민’ 등으로 나타났다. 추가로 ‘다리’, ‘하천’, ‘공사’, ‘장소’ 등 발견 장소를 특징하는 키워드가 네트워크상에서 확인되었다. 특히, 멸종위기 야생생물에 대한 정보, 자문 등 요청이 다양한 주체를 통해 이루어지는 것을 유추할 수 있다. 이러한 토픽으로 분류된 민원 데이터 원본 내용을 살펴보면 단체, 기자 등의 멸종위기 야생생물 발견과 연관 질문, 인터뷰 요청, 정보 제공 요청 등에 대한 민원이 많았다.

IV. 결론

본 연구는 멸종위기 야생생물 민원 데이터를 수집·분석하여 국민의 수요와 관심을 파악하고,

멸종위기 야생생물 관련 주요 이슈와 토픽을 도출하였다. 1,739건의 민원 텍스트 데이터가 8개의 토픽과 5개의 클러스터로 분류되었으며, 각 토픽 별 키워드간의 관계를 해석하였다. 이를 바탕으로 다음과 같은 시사점과 정책 활용방안을 제안한다.

국민은 생활권에서 발견되는 멸종위기 야생생물에 대해 관심이 있으며 대중매체를 통해 널리 알려져 인지도가 높은 맹꽁이, 수원청개구리와 같은 개발사업 관련 이슈중, 수달, 삶, 담비와 같은 중소형 포유류들에 대한 민원이 많았다. 또한, 발견 제보 민원을 통해 멸종위기 야생생물 보전에 기여하고자 하며 특히 주거지 인근의 서식지 보호와 관리를 요구하는 민원이 많았다.

곤충, 조류, 육상식물 등 일부 분류군의 경우, 일회성 발견 제보뿐만 아니라 추가 중 서식지 확인 요청이나 지속 관찰 결과 공유 등 신규 서식지 발굴과 모니터링에 적극적인 사례가 다수 있었다. 이러한 지역 기반의 국민 참여를 보전 활동과 연계하여, 상세 위치와 생태환경을 포함하는 지속가능한 멸종위기 야생생물 서식지 정보 구축을 위한 국민-지역 참여형 멸종위기 야생생물 모니터링 체계를 운영하는 것이 필요하다.

연도별 주요 키워드 변화와 토픽 네트워크 분석을 통해, 멸종위기 야생생물에 대한 질의가 다양화되고 전문화되고 있음을 알 수 있었다. 국가적색목록이나 국제적멸종위기종, 멸종위기 야생생물 포획·채취 등의 법률적 문제부터 특정 종의 생태나 서식지 등의 전문적 문제까지 범위가 확장되고 있어, 이에 대응하기 위한 민원 서비스 체계 개선이 시급하다.

소똥구리에 대한 민원은 단일 멸종위기종 민원 중 가장 많은 건수였다. 이는 초기 복원 사업에서 자생 소똥구리 확보를 위한 공고가 현상금 지급 형태로 오해되어, 현상금 목적으로 곤충을 포획 후 발견, 제보 민원으로 건의하는 사례가 많았기 때문이다. 정정 보도와 홍보가 이루어진 바 있으나, 여전히 소똥구리 동정 확인 요청과

현상급 지급 불가에 대한 부정적 반응이 많다. 이에 대한 효과적인 대응 방안이 필요하다. 맞춤형 정보를 시각화하여 제공하고, 피드백을 받는 기반을 마련하여, 국민의 관심과 필요에 따른 정확한 정보를 제공하는 것이 하나의 대응 방안이 될 수 있다.

멸종위기 야생생물 보호를 위한 조사 데이터 공유 요청 민원이 많았다. 현재 멸종위기 야생생물 데이터 관련하여, 위치 정보를 공유하지 않는 것을 원칙으로 하고 있었으나, 본 연구를 통해 멸종위기 야생생물, 특히 서식 정보에 관한 관심이 증가하고 있음을 알 수 있었다. 따라서 멸종위기 야생생물의 보호와 국민의 관심과 수요 충족을 위한 다양한 서비스 제공이 필요할 것으로 보인다. 예를 들어 시군구 단위의 위치 정보 갱신 주기 조절, Web-GIS 형태로 정보 제공하는 등과 같은 지속가능한 정보 공유 등의 노력이 요구된다. 이를 통해 신규 서식지 발굴 등에도 긍정적 영향을 줄 수 있으며, 국민의 보전 활동 참여 의지도 높일 수 있을 것이다. 또한, 멸종위기 야생생물 관련 정보 공유와 생산을 할 수 있는 빅데이터 플랫폼을 구축하여, 모두가 함께 즐기고 참여하는 양질의 경험 가치를 제공하는 것을 목표로 할 수 있다. 이러한 플랫폼을 통해 멸종위기 야생생물 관련 빅데이터 수집과 분석도 더욱 효율적으로 할 수 있을 것이다.

본 연구는 멸종위기 야생생물 관련 정책 의사결정의 기초 자료로 활용될 수 있으며, 민원 데이터 분석을 통해 시사점을 제시하고 멸종위기 야생생물 관련 빅데이터 활용을 도모하였다. 그러나 본 연구에는 민원인 정보 수집 및 활용 불가, 한정된 창구의 민원 데이터 수집, 공간적 분석 불가 등의 한계가 있었다. 따라서 향후 본 연구를 기초 자료로 하여 데이터 추가 수집, 멸종위기 야생생물 정책 관련 수요와 공급자에 대한 데이터 분석 등의 후속 연구를 수행한다면, 국민의 관심과 수요를 반영한 멸종위기 야생생물 관련 정책 의사결정 자료로 더욱 폭넓게 활용될

수 있을 것으로 기대된다.

References

- Blei, D.M. 2012. Probabilistic Topic Models. *Communications of the ACM*, 55, 77-84. <http://dx.doi.org/10.1145/2133806.2133826>
- Choi CM, Lee JH, Jessica Machado, and Kim GW. 2022. Big-Data-Based Text Mining and Social Network Analysis of Landscape Response to Future Environmental Change *Land* 11, no. 12: 2183. <https://doi.org/10.3390/land11122183>
- Choi HO. 2016. Study on Selecting Priority Criteria Utilizing Civil Complaint Data in the Field of Environment and Sanitation. *Journal of Environmental Policy and Administration*, 24(2), 45-57.
- Chae HH, Kim YC, Son SW. 2022. Korean and Worldwide Research Trends on Rare Plant and Endemic Plant in Korea. *Korean J. Environ. Ecol.* 36(3): 257-276, June 2022
- Chung JM, Kim WJ, Koo CD. 2016. Social Media Big Data Analysis for ICT Policy Agenda in Education. *Korea Education and Research Service(KERIS)*, pp.4
- CIVIL PETITIONS TREATMENT ACT article No.2 [Act No. 18742, Jan. 11, 2022]
- Do MS, Choi GR, Hwang JW, Lee JY, Hur LH, Choi YS, Son SJ, Kwon IK, Yoo SY, Nam HK. 2020. Research topics and trends of endangered species using text mining in Korea. *Journal of Asia-Pacific Biodiversity*, 13(4), 2020, 518-523. ISSN 2287-884X,
- Duberry, J. 2019. Big data and environmental civil society organizations. In *Global Environmental Governance in the Information Age*:
- Jiao Y, Li C, and Lin Y. 2021. Can Urban Environmental Problems Be Accurately

- Identified? A Complaint Text Mining Method. *Applied Sciences*, 11(9):4087. <https://doi.org/10.3390/app11094087>
- Jin DY, Kang SW, Han KJ, Kim JH, Kim DY and Kang SA. 2019. A Study on Improving Reflection of Demands for Settling Environmental Issues in Daily Lives: Focusing on the Analysis of Civil Complaint Big Data. Korea Environment Institute, 2019-09
- Lee HJ & Sung KJ. 2018. Analysis of Domestic and Foreign Local Biodiversity Strategies and Action Plan (LBSAP) using Semantic Network Analysis. *Journal of Environmental Impact Assessment*, 27(1), 92-104.
- Lee SY. 2021. An Analysis of Civil Complaints about Traffic Policing Using the LDA Model. *The Journal of The Korea Institute of Intelligent Transport Systems*, 20(4), 57-70.
- Lim HI, Park GW, and Kim AR. 2023. Challenges in the Forestry Sector for Achieving the Kunming-Montreal Global Biodiversity Framework, a New Leap Forward for Global Biodiversity Conservation. *International forest agenda vol.123*. National Institute of Forest Science
- Lim HJ. 2014. Analysis of Chungnam Province Policy Keywords Using Big Data. Chungnam Development Institute
- Ministry of Environment [website]. (2022.12.20.). URL: <http://www.me.go.kr/home/web/board/read.do?boardMasterId=1&boardId=1569450&menuId=10525>
- Ministry of Environment. 2023. National Biodiversity Strategy(2024~2028)
- National Institute of Ecology. 2020. National Survey on Public Awareness of Endangered Wildlife-1st
- National Institute of Ecology. 2022. National Survey on Public Awareness of Endangered Wildlife-2nd
- Park JS & Lee SM. 2020. Big Data Analysis of Busan Civil Affairs Using the LDA Topic Modeling Technique. *Informatization policy*, 27(2), 66-83.
- Park KC. 2020. Big data analysis of complaints using text mining techniques: Gangnam-gu. Seoul: Seoul Digital Foundation
- Ryu SE, HONG SG, Lee TH, KIM NR. 2018. A Pattern Analysis of Bus Civil Complaint in Busan City Using the Text Network Analysis. *Korean Computers and Accounting Review*, 16(2), 19-43.
- Secretariat of the Convention on Biological Diversity. 2020. Global biodiversity Outlook 5. Montreal
- Sung JH & Lee KJ. 2022. An Analysis of Citizen's Complaints in Urban Parks Using Text Mining. *Journal of recreation and landscape*, 16(4), 99-104
- Wallach, Hanna & Murray, Iain & Salakhutdinov, Ruslan & Mimno, David. 2009. Evaluation methods for topic models. *Proceedings of the 26th International Conference On Machine Learning, ICML 2009*. 382. 139. <https://doi.org/10.1145/1553374.1553515>.
- Yang HJ, Ahn JM, and Lee TH. 2021. A Study of Korean's Experiences of Unfairness Based on Analysis of Text Big Data Posted on the Blue House National Petition. *Survey Research*, 22(1), 25-59.
- Yoon J., Kwon K., Yoo, J., and Yoo, N. 2021. Current Status and Future Prospects of Endangered Species Restoration Projects for Freshwater Fishes, Amphibians, and Reptiles in South Korea. *Proceedings of the National Institute of Ecology of the Republic of Korea*, 2(4), 247-258, <https://doi.org/10.22920/PNIE.2021.2.4.24>