

텍스트마이닝을 활용한 도로분야 ITS 정책이슈 탐색기법 정립

Establishment of ITS Policy Issues Investigation Method in the Road Section applied Textmining

오 창 석*
(Chang-Seok Oh)
(The Board of Audit and
Inspection of Korea)

이 용 택**
(Yong-taeck Lee)
(The Board of Audit and
Inspection of Korea)

고 민 수***
(Minsu Ko)
(Korea Advanced Institute of
Science and Technology)

요 약

본 연구는 빅데이터를 활용하여 감사 시 유의해서 살펴보아야 할 ITS 관련 정책이슈 탐색방법 개발 및 적용을 목적으로 한다. 이를 위해 본 연구에서는 William Dunn이 제안한 경계분석을 이론적 토대로 하여, 여기에 감사원 감사실무 프로세스를 접목한 감사이슈 분석 틀을 제안했다. 그리고 이 분석 틀을 전산으로 구현하기 위해 메타문제를 추정하는 개념이 경계분석과 유사한 텍스트마이닝 기법을 응용했다. 텍스트마이닝의 구체적 모형은 David Blei가 제안한 Latent Dirichlet Allocation(LDA) 모형을 기반으로 하는 비대칭-대칭 혼합 어휘소 기반 LDA를 응용했다. 사례분석 결과, 경찰청에서 운영하는 도시교통정보시스템의 교통정보 수집률 저조와 국토교통부의 첨단교통관리시스템과의 중복 문제, 디지털 운행기록계의 주행거리 조작 등이 주요 이슈로 도출됐다.

핵심어 : 빅데이터, 경계분석, 텍스트마이닝, 주제분석, LDA 모형

ABSTRACT

With requiring circumspections using big data, this study attempts to develop and apply the search method for audit issues relating to the ITS policy or program. For the foregoing, the auditing process of the board of audit and inspection was converged with the theoretical frame of boundary analysis proposed by William Dunn as an analysis tool for audit issues. Moreover, we apply the text mining technique in order to computerize the analysis tool, which is similar to the boundary analysis in the concept of approaching meta-problems. For the text mining analysis, specific model we applied the antisymmetry-symmetry compound lexeme-based LDA model based on the Latent Dirichlet Allocation(LDA) methodologies proposed by David Blei. The several prime issues were founded through a case analysis as follows: lack of collection of traffic information by the urban traffic information system, which is operated by the National Police Agency, the overlapping problems between the Ministry of Land, Infrastructure and Transport and the Advanced Traffic Management System and fabrication of the mileage on digital tachograph.

Key words : Big Data, Boundary Analysis, Text Mining, Topic Analysis, LDA Model

† 이 논문은 [1]Oh(2016), 『빅데이터와 텍스트마이닝을 활용한 감사이슈 도출기법 연구』를 금번 학회지에 맞도록 수정·보완하여 게재한 것임을 밝힌다.

* 주저자 : 감사원 감사연구원 연구관

** 교신저자 : 감사원 행정·안전감사국 제1과 수석감사관

*** 공저자 : 한국과학기술원 전산학과 연구원

† Corresponding author : Yong-taeck Lee(The Board of Audit and Inspection of Korea), E-mail icarus01@korea.kr

† Received 20 September 2016; reviewed 20 October 2016; Accepted 9 November 2016

I. 서론

소위 ‘빅데이터(Big Data)’로 일컬어지는 대규모의 사회적 데이터는 컴퓨터와 웹의 발달로 최근 그 활용범위가 확대되는 추세이다. 이런 흐름을 반영하듯 경제협력개발기구(OECD)는 빅데이터가 정부의 정책형성 과정을 지원할 수 있는 잠재능력을 가지고 있으며, 정부와 국민들 간의 신뢰를 향상시킬 수 있는 정책수단이 될 것으로 전망하고 있다[2].

특히, EU와 미국 등 선진국에서는 최근 공공감사행정에 위험기반 감사접근방법의 일환으로 대용량 데이터를 활용한 마이닝기법의 활용을 적극적으로 도입하는 등 빅데이터를 이용한 감사이슈 탐색 방법론을 구축·활용하고자 노력하고 있다. 실제로, 미국 연방감사원(U. S. Government Accountability Office, GAO)은 2015년 공공부문의 다양한 공개정보를 활용한 감사자료 분석기법을 개발하고, 이를 실제 감사행정에 활용할 수 있도록 테스트 중에 있으며[3], ECOPA(2014)는 감사유형에 따른 세분화된 마이닝 기법을 활용할 수 있는 감사계획 설계방법을 연구하였다[4].

본 연구도 이상에서 살펴 본 빅데이터를 활용한 감사방법론 개발 시류에 일조하고자 했다. 연구의 목적은 빅데이터를 활용하여 공공감사 시 유의해서 살펴보아야 할 정책이슈 탐색방법을 개발하고 적용함에 있다. 먼저 방법론 개발을 위해 국내 인터넷 환경에 널리 퍼져있는 텍스트 형태의 비정형적 빅데이터로부터 공공정책의 이슈를 간편하게 탐색할 수 있도록 경계분석 이론과 텍스트마이닝(Text Mining)을 활용한 정책이슈 탐색기법을 정립하고자 한다. 그리고 정립된 방법론을 토대로 도로부문 지능형교통체계(ITS)를 대상으로 사례적용을 함으로써 방법론의 적용가능성도 테스트해보고자 한다.

II. 이론적 고찰

본 연구의 정책이슈 탐색 Dunn(2008, 2012)의 경계분석(Boundary Analysis)에 이론적 토대를 두었다 [5, 6]. 그리고 경계분석과 메타문제를 추정하는 개

념이 유사한 텍스트마이닝을 활용하여 빅데이터 기반 정책이슈 탐색방법론을 정립했다. 본 절에서는 이론적 배경을 이루는 경계분석 이론과 텍스트마이닝 기법에 대하여 고찰한다.

1. 경계분석 이론에 대한 이론적 고찰

Dunn(2008, 2012)은 정책이슈를 구조화하는 방법으로 경계분석(boundary analysis)을 제안하였는데, 이는 정책에 대한 이해관계자의 의견을 종합적으로 수렴하기 위해 사용하는 정책의사 결정방법으로 널리 활용되고 있다[5, 6]. 또한 경계분석 과정은 일반적으로 ‘포화표본 추출(saturation sampling)’, ‘문제의 대표성 도출(elicitation of problem representation)’, ‘경계추정(boundary estimation)’의 3단계로 구분한다 [5, 6].

첫째, ‘포화표본 추출’ 단계는 정책에 대해 이해관계자들이 요구하는 정책이슈가 쌓여서 이론 ‘눈덩이(snowball)’를 포화표본(saturation sample)으로 추출하는 과정이다. 이 과정은 해당 이슈에 대하여 더 이상의 새로운 이해관계자가 나타나지 않을 때까지 반복하여 추출한다[6]. 이 과정을 전산으로 구현하기 위해 텍스트마이닝 분석을 위한 자료 분류체계(taxonomy)를 구축하고, 이 분류체계를 토대로 빅데이터 탐색기법인 크롤링(crawling) 기법을 주로 활용하고 있다.

둘째, ‘문제의 대표성 도출’ 단계는 첫째 단계에서 추출된 포화표본을 주제별로 해석하는 과정으로 이해관계자들이 인식하는 집단화된 정책이슈를 이끌어내는 과정이다[6]. 이 과정을 전산으로 구현하기 위해 텍스트마이닝 기법의 일종인 ‘주제분석(topic analysis)’이 널리 활용된다.

셋째, ‘경계추정’ 단계는 다양한 이해관계자로부터 의견을 수렴하고 이를 정량화하기 위하여 메타문제의 경계를 추정하는 과정이다[6]. 예를 들어 경계추정을 위하여 2가지 주요 지표로 플랫폼별 노출빈도(exposure frequency)와 정책이슈의 주제비중(topic weight)을 선정하였다고 가정하면, 먼저 수평축에 정책대상에 대한 플랫폼별 노출빈도(문서의 경우 발간일자)를 나열하고, 수직축에 문서의 누적

출현 건수를 산출하여 정책대상의 플랫폼별 노출빈도를 판단한다. 그리고 ‘주제비중’은 텍스트마이닝 분석을 통해 나타나는 모형의 추정 값으로 산출되는 관련정보의 유사주제 집합에 대해 ‘집합별-문서별 주제비중’을 추정한 후 종합적 판단을 통해 정책이슈를 선정하게 된다.

2. 텍스트마이닝에 대한 이론적 고찰

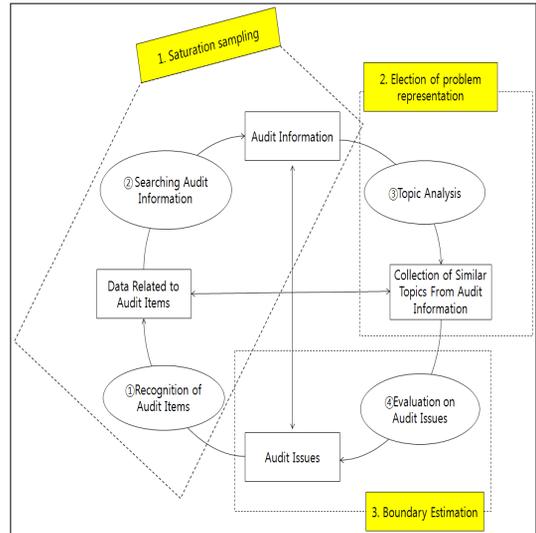
Hearst(1999, 2003)[7, 8]는 텍스트마이닝을 “다르게 작성된 정보 원천으로부터 예전에 알려지지 않은 정보를 추출하여 컴퓨팅을 통해 새로운 정보를 발견하는 것”으로 정의하였다[9]. 텍스트마이닝은 데이터마이닝(data mining)의 일종으로, 텍스트 형태로 이루어진 비정형 데이터들을 자연어 처리방식을 이용해 정보를 추출하는 방법이다. 이는 수치형 데이터(numerical data)나 범주형 데이터(categorical data)로부터 의미 있는 패턴을 찾아내는 데이터마이닝에 비해 텍스트 데이터로 정보를 다룬다는 점이 특징이다[8, 10]. 따라서 텍스트마이닝의 기술적 정의는 비정형 또는 반정형 데이터에 대하여 자연어 처리(Natural Language Process, NLP) 기술과 문서처리 기술을 적용하여 유용한 새로운 정보를 추출, 가공하는 정보처리기법이라 할 수 있겠다[11]. 일반적으로 텍스트마이닝의 프로세스는 ‘비정형 정보수집 → 전 처리 → 정보추출 → 정보분석’의 일반적인 절차를 따르고 있으며, 정보추출과정에서 수학적 모델이나 알고리즘을 통해 유용한 정보를 추출한다[11].

III. 정책이슈 탐색방법론 정립

본 연구에서는 Dunn(2012)가 제안한 경계분석을 이론적 토대로 하여, 여기에 감사원 감사실무 프로세스를 접목한 감사이슈 분석 틀(analogical framework)을 <Fig. 1>과 같이 제안한다.

<Fig. 1>에서 보는 바와 같이 본 연구에서 제안하는 경계분석 절차는 Dunn(2012)의 제안대로 ‘1.포화표본 추출(saturation sampling)’, ‘2.포화표본 추출(elicitation of problem representation)’, ‘3.경계추정

(boundary estimation)’으로 구성했다. 그리고 그 안에 배치되는 텍스트마이닝 분석요소는 ‘① 정책대상 인식(recognition of audit items)’, ‘② 정책정보 탐색(searching audit information)’, ‘③ 주제분석(topic analysis)’, ‘④ 정책이슈 판단(evaluation on audit issues)’이라는 4단계로 이루어진 감사이슈 구조화 과정으로 이루어져 있다.

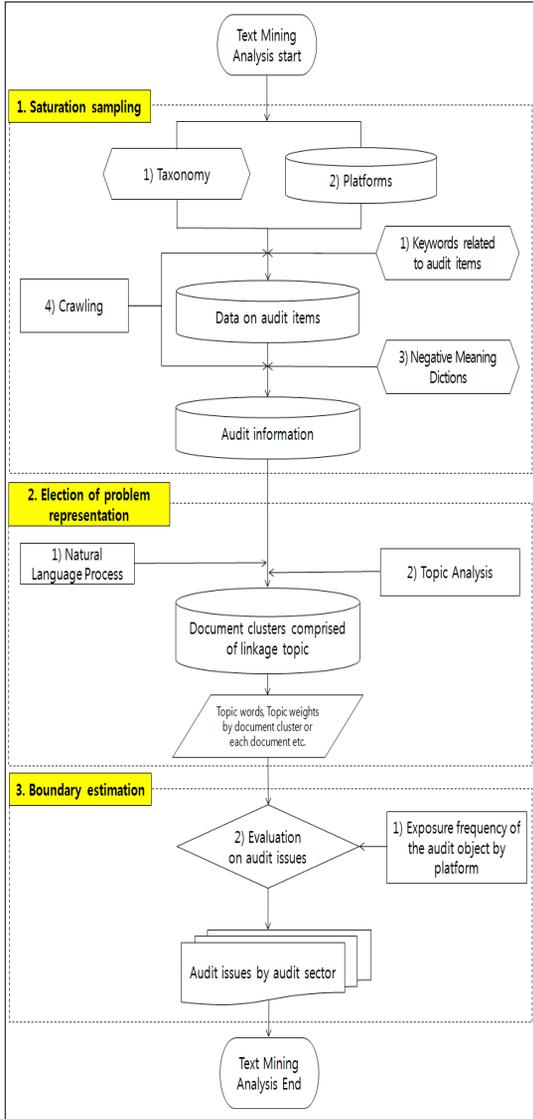


<Fig. 1> Analogical boundary analysis framework applied Text Mining for audit issues investigation

Annotation 1: The ellipses indicate input activities or analysis activities for analyzing the audit issues, the rectangles indicate output activities or audit activities and the shaded rectangles and dotted lines indicate the areas of the analysis tool that match with the boundary analysis proposed by Dunn (2012).

Annotation 2: The brackets indicate the text mining techniques (①, ②, ③) applied in the study or the values estimated therefrom (④).

이상에서 제안한 감사이슈 분석 틀을 빅데이터 기법인 텍스트마이닝을 활용하여 구현하고자 하는데 그 방법을 분석 단계별로 세부적으로 살펴보면 <Fig. 2>와 같다.



<Fig. 2> The boundary analysis flow chart for audit issues investigation using Text Mining

Annotation: Analysis items in each stage were consistent with the audit activities shown <Fig. 1> and dashed boxes display the boundary analysis proposed by Dunn(2012).

첫째, ‘정책대상 인식’ 과정은 정책대상을 체계적으로 인식하여 관련 자료를 추출하기 위한 과정이다. 즉, 구체적으로 정책대상을 구체적으로 인지하고, 이와 관련한 자료를 ‘어디에서 찾을지’를 해결하기 위

한 과정이다. 이를 위해 빅데이터 분석에서 많이 사용하는 분류체계를 구성하고, 텍스트마이닝 기법을 정책이슈 탐색을 위한 ‘플랫폼(platform)’으로 활용했다.

둘째, ‘정책정보 탐색’ 과정은 정책대상 관련 자료 중 의미 있는 정책정보를 찾기 위한 과정이다. 즉, 첫 번째 단계인 ‘정책대상 인식’ 단계에서 구성된 분류체계와 플랫폼을 이용하여 정책정보를 실질적으로 탐색하는 단계이다. 본 연구는 효율적인 탐색을 위하여 텍스트마이닝 자료추출기법인 ‘크롤링’을 활용했다. 이는 온라인이나 오프라인상에서 검색 키워드를 활용하여 관련 자료를 수집하고, 수집한 자료를 텍스트 정보로 변환하는 기능을 프로그램 코딩을 통하여 컴퓨터가 자동으로 처리하도록 하는 기법이다.

본 연구에서는 ‘.NET’과 ‘MS-ACCESS Visual Basic’을 통해 플랫폼별 검색구조에 적합한 크롤러를 코딩했다. 이상의 첫 번째 ‘정책대상 인식’ 과정과 두 번째 ‘정책정보 탐색’ 과정은 <Fig. 2>의 ‘1. 포화표본 추출’에 해당한다.

셋째, ‘주제분석’ 과정은 두 번째 단계에서 탐색 및 선별된 정책정보를 유사한 주제를 갖는 정보끼리 묶어 유사주제 집합으로 만드는 과정이다. 이는 Dunn(2012)의 경계분석 과정 중 ‘문제의 대표성을 도출’하는 과정이라고 할 수 있다. 즉 주제분석은 정책정보에서 유사한 주제를 갖는 정보끼리 묶고, 이를 대표하는 정보가 무엇인지를 파악하는 과정이다.

본 연구는 이를 빅데이터기반 분석기법으로 전환 (conversion)하기 위하여 자연어 처리(Natural Language Process, NLP)기술과 LDA(Latent Dirichlet Allocation) 알고리즘을 적용한 주제분석을 수행했다. 자연어 처리기술은 인간이 사용하는 일상적인 언어인 자연어 (natural language)를 컴퓨터가 인식할 수 있는 의미 단위로 데이터의 기준을 만들어주는 과정으로 빅데이터 정제과정이라고도 할 수 있다.

또한, 주제분석(topic analysis)은 정책대상 관련 자료에서 추출된 정보를 유사주제를 갖는 문서끼리 집합화(clustering)하는 기법이다. 주제분석 모형은 LSA (Latent Semantic Analysis) 모형, PLSA(Probabilistic

Latent Semantic Analysis) 모형, LDA(Latent Dirichlet Allocation) 모형 등이 있으며[11], 본 연구에서는 수집된 문서들 간의 상호의존성(inter-dependency)과 정책이슈의 포괄성(diversity)을 잘 반영하는 ‘LDA 모형’을 사용했다.

여기서 LDA 모형은 문서들의 주제를 분류하기 위해 Blei et al.(2003)가 개발한 모형으로[14], 발견되지 않은 그룹에 의해 설명될 수 있는 관측치의 집합을 설명하는 생성모델(generative model)이라 할 수 있다. 즉, LDA 모형은 잠재적 확률추정기법을 사용하며, 이 기법은 하나의 문서는 여러 개의 주제를 포함하거나 여러 문서가 공통의 주제를 공유할 수도 있다는 점을 전제로 하고 있으며, 각 주제는 일정한 분포를 가지고 있고, 각 주제의 분포는 데이터 전체 분포의 하위집합으로 구성되는 위계모형(hierarchical model)의 특성을 지닌다.

주제분석의 결과로는 유사주제 집합을 대표하는 주제어(topic words)가 산출되고, 문서군별 주제비중(topic weights by document cluster), 문서별 주제비중(topic weight by each document)이 추정되는데, 이는 <Fig. 2>의 ‘2. 문제의 대표성 도출’에 해당한다.

셋째, ‘정책이슈 판단’은 주제분석을 통해 추정된 정책정보의 유사주제 집합을 토대로 정책이슈를 정성적으로 판단하는 과정이다. 이는 경계분석에서 경계추정에 해당하는 과정으로 <Fig. 2>의 ‘3. 경계 추정’에 해당한다.

앞의 세 과정은 컴퓨터에 의존하였다면, 네번째 과정은 정책이슈를 분석하는 분석가의 인적 판단과정이라고 할 수 있다. 본 연구는 정책이슈 판단을 위해 감사대상의 노출빈도(audit object's exposure frequency by platform)와 주제분석 결과 도출되는 주제어와 문서군별 주제비중 그리고 문서별 주제비중 등을 판단지표로 활용했다. 또한 소개한 모든 과정은 순환구조를 갖게 되며, 분석가는 일정한 정책이슈를 도출하기까지 이를 반복 수행할 수 있다.

IV. 사례 적용

본장에서는 3장에서 정립한 정책이슈 탐색기

법을 감사행정에 활용할 수 있도록 테스트하고자 ITS 정책이슈를 대상으로 사례적용을 수행했다. 이에 <Fig. 2>에서 제시한 텍스트마이닝을 활용한 감사이슈 경계분석 절차에 따라 각 단계별로 정책이슈를 탐색한 방법과 결과를 서술하면 다음과 같다.

1. 포화표본 추출(Saturation sampling) 단계

1) 분류체계(Taxonomy) 및 감사대상 관련 키워드(Keywords related to audit items) 설정

‘분류체계 수립’은 ‘무엇을 감사할지’를 해결하기 위한 과정으로 감사대상 관련 자료를 효과적으로 추출하기 위한 과정으로, 빅데이터 플랫폼은 다양한 이해관계자가 생산한 방대한 양의 자료를 포함하고 있기 때문에 이 중에서 감사대상과 유관한 정보만을 효과적으로 추출하기 위해서는 관련 자료에 대한 제한이 필요하다. 따라서 본 연구에서는 국제기준(INTOSAI IT감사 분류기준)을 참고하여 ‘감사분야 - 분야별 감사대상 - 감사대상 관련 키워드’로 이루어진 분류체계를 구성했다. 본 연구에서 선정한 감사대상과 관련한 분류체계 및 키워드는 <Table 1>과 같이 2개 감사분야, 14개 감사대상, 62개 감사대상 관련 키워드로 구성했다. ‘감사대상 관련 키워드’는 다음 단계에서 진행될 플랫폼별 크롤링 과정의 입력변수로 사용했다.

2) 플랫폼(Platforms) 구성

‘플랫폼 구성’은 분류체계를 토대로 ‘관련 감사자료를 어디에서 찾을지(Where will I audit?)’를 선정하는 과정으로 앞서 수립한 분류체계의 키워드 집합을 가지고 여러 이해관계자로 구성된 플랫폼을 통해 논의되고 있는 다양한 정보를 탐색한다.

본 연구는 플랫폼을 대중 플랫폼과 전문 플랫폼으로 구분했다. 대중 플랫폼은 웹상에서 누구나 접근 가능한 범용 인터넷 포털인 ‘네이버’, 뉴스 포털과 ‘다음’ 뉴스 포털을 이용했다. 그리고 대형포털의 풀립현상을 보완하고 지방지역의 여론을 균형 있게 수집하기 위해 ‘한국언론재단’에서 제공하는

<Table 1> Group of keywords about the audit items and taxonomy thereof in the study

Areas of Audit	Audit items in each section	Keyword related to the audit items
Develop-ment or Establish-ment	1-① CCTV Parking Enforcement System	CCTV parking enforcement system, parking enforcement system, CCTV parking enforcement, CCTV, parking violation enforcement, closed circuit parking enforcement system, parking enforcement system closed circuit
	1-② Advanced Traffic Management System (ATMS)	ATMS, Advanced Traffic Management System
	1-③ Digital Tachograph (DTG)	DTG, Digital Tacho Graph, Digital tachograph, digital tachograph device, digital tachometer, digital tacho, digital driving, digital driving device
	1-④ Bus Information System (BIS)	BIS, bus information system, bus guidance information system, bus arrival information system, bus operation information system
	1-⑤ Cooperative Intelligent Transport Systems (C-ITS)	C-ITS, Cooperative ITS, cooperative intelligent transport systems, V2V, V2I, Vehicle to Vehicle, Vehicle to Infrastructure, Vehicle to Infra, Vehicle to Person
	1-⑥ Traffic Accident Analysis System (TAAS)	TAAS, traffic accident analysis system, traffic accident analysis DB, the national police department traffic accident analysis system, the national police department traffic accident analysis DB
	1-⑦ Electronic Toll Collection System (ETCS)	ETCS, electronic toll collection system, automatic toll collection system, electronic toll payment system
	1-⑧ Traffic Safety Information Management Complex System (TMACS)	TMACS, traffic safety information management system
	1-⑨ Urban Traffic Information System (UTIS)	UTIS, urban traffic information system, central traffic information center
Operation and Management	2-① Traffic Information Connections	traffic information communication, traffic information connections, traffic information sharing
	2-② Traffic Management Center	traffic management center, traffic information center, traffic information management, traffic information management center, local autonomous community traffic information management center
	2-③ Traffic Information Overlaps	traffic information overlap, overlap of traffic information
	2-④ Traffic Information Application	Traffic information application, application of traffic information, traffic information usage
	2-⑤ T-map and other traffic apps	T-MAP, real-time traffic information apps(or applications), public transportation apps(or applications), traffic apps(or applications)

온라인 뉴스 서비스인 「미디어가온」의 「뉴스기사」 서비스 중 「지역종합일간신문」, 「인터넷·전문신문」, 「지역주간신문」, 「계시판 등(이하, 「미디어가온」 지방지 포털) 205개 지방언론 매체를 소스로 활용했다. 자료 수집의 시간적 범위는 2012년 1월 1일에서 2014년 12월 31일까지 3년간으로 설정했다.

3) 부정적 견해 어휘(Negative Meaning Dictions) 설정

언론매체 및 전문 문헌은 긍정적, 중립적, 부정적

견해를 갖고 있는데, 감사에 필요한 정책이슈를 탐색하기 위해서는 ‘플랫폼’을 대상으로 ‘분류체계’와 ‘감사키워드’를 통해 부정적 견해를 갖는 감사대상 관련 자료를 선별하여 관련 정보를 추출하는 과정이 필요하다.

본 연구에서 사용한 부정적 견해어 집합은 <Attachment 1>과 같은데, 기존 감사원 감사보고서와 대중언론 보도 자료에서 빈출되게 사용되고 있다고 판단되는 어휘 270개를 정리했다. 그리고 이

부정적 견해어를 포함하고 있는 문서를 크롤러를 활용하여 문장 단위로 인터넷상에서 탐색한다. 탐색방식은 보도 자료(또는 문서)의 제목에 <Table 1>에서 제시한 감사키워드와 부정적 견해어를 동시에 포함하고 있는 문서를 추출하거나 또는 본문에서 문장 단위로 탐색하여 감사키워드와 부정적 견해어를 동시에 포함하는 문서를 ‘정책이슈’로 추출했다.

4) 크롤링(Crawling) 과정

‘크롤링’ 과정은 검색 키워드를 활용하여 온라인 등에서 해당 자료를 수집하고, 수집한 자료를 텍스트 정보로 변환하는 기능을 컴퓨터가 자동으로 처리하도록 하는 과정이다. 본 연구에서 사용한 크롤링 기법은 <Table 2>와 같이 방법적 분류에 따라 ‘웹 대상 크롤링(web targeted crawling)’ 방식과 ‘파일 대상 크롤링(file targeted crawling)’ 방식으로 구분되며, 기능적 분류에 따라 ‘수집기능 크롤러(collector targeted crawling)’와 ‘저장기능 크롤러(saver targeted crawling)’로 구분된다.

방법적 분류 기준에 따른 크롤링 기법을 살펴보면, 먼저 ‘웹 대상 크롤링’ 방식은 인터넷 포털뉴스 검색란에 관련 키워드를 입력하고, 이때 나타나는 검색결과 중 원하는 링크를 클릭하여 열리는 웹 문서를 읽는 인간의 탐색방식을 자동화하여 구현한 방법이다. 본 연구에서는 중앙미디어, 지방미디어, 전문미디어 등 웹에서 링크를 통해 열람할 수 있는 보도 자료 탐색에 주로 활용했다. 다음으로 ‘파일 대상 크롤링’ 방식은 관련 자료가 웹 문서로 이루어지지 않고 pdf 파일 등의 형태로 되어 있는 경우 관련 자료를 다운로드한 후 파일을 매체로 크롤링하는 방법이다. 본 연구에서는 국회, 국책연구기관, 관련 학회 등 홈페이지에서 ‘감사대상 관련 키워드’를 이용해 감사대상 관련 자료를 추출하는 방식을 이용했다.

기능적 분류 기준에 따른 크롤링 방식을 살펴보면, ‘수집기능 크롤러’는 웹이나 다운로드한 문서에서 분류 체계에 해당하는 키워드를 가지고 일정 조건에 부합하는 정보를 수집하는 방법이다. 본 연구에서는 플랫폼별로 정보검색 구조를 분석하여 ‘.NET’으로 코딩

했다. 또한, ‘저장기능 크롤러’는 수집기능 크롤러를 통해 수집된 감사대상 관련 자료 및 감사정보를 플랫폼별, 분류체계별 등 특정 색인으로 구분하는 방법이다. 본 연구에서는 ‘MS-ACCESS Visual Basic’을 통해 코딩하여 수집된 관련 자료 및 정보를 로컬 PC 내에 ACCESS DB 형태로 저장할 수 있도록 했다.

<Table 2> Crawling technic implemented in this study

Classification	Using Technic	Description of Functions	Coding Language
Methodological classification	Web targeted crawling	Crawling method applied for searching report materials such as central media, local media and professional media, which can be viewed through online links	.NET
	File targeted crawling	Crawling method used when materials, like those from the national assembly, national research institutes and relevant colloquiums, exist in file links instead of web documents, to download relevant documents and crawling the files.	.NET
Functional classification	Collector targeted crawling	Method used to explore web or text documents using key words and negative comments, set in the taxonomy	.NET
	Saver targeted crawling	Method used to save text documents collected by ‘collector-targeted crawler’ on PC	MS-ACCESS VB

2. 문제의 대표성 도출(Elicitation of problem representation) 단계

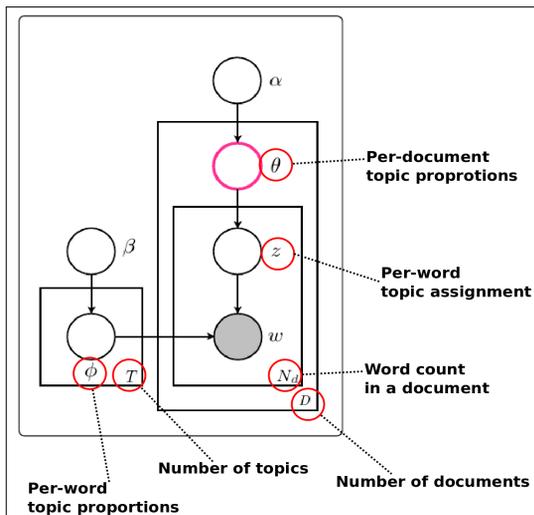
1) 자연어 처리(Natural Language Processing) 과정

‘자연어 처리 과정’에서는 크롤링 된 자연어(natural language) 형태의 자료 또는 정보를 컴퓨터가 인식할 수 있는 의미단위로 데이터의 기준을 부여한다. 본 연구에서는 띄어쓰기 처리, 오타 교정, 동의미 이형태 처리, 형태소 분석, 개체명 인식, 구문 분석 등 총 6가지의 과정을 거쳤다.

2) 주제분석(Topic Analysis) 과정

‘주제분석 과정’은 앞의 <Table 1>에서 설정한 감사대상 관련 키워드를 입력하여 크롤링한 문서들을 텍스트마이닝 모형을 적용하여, 유사한 주제를 갖는 감사정보끼리 클러스터링(집합화)하는 단계라 할 수 있다. 주제분석을 통해서 추정되는 값은 ‘주제어’와 ‘주제비중’이며, 이는 다음 단계의 감사이슈 판단에 활용된다. 이는 크롤링을 통해 수집한 감사대상 관련 문서를 자연어 처리를 통해 키워드를 산출하고, 이 키워드를 토대로 유사한 주제를 집합화하는 과정이다.

텍스트마이닝 과정에서 ‘주제분석’은 문서의 주제를 알기 위해 원본 텍스트 내의 단어를 분석하는 통계적인 방법으로, 인공지능 학습모형에 기반하여 대량의 문서에서 원하는 주제를 빠르게 도출하는 역할을 한다[12, 13]. 주제분석 모형에는 LSA(Latent Semantic Analysis) 모형, PLSA(Probabilistic Latent Semantic Analysis) 모형, LDA(Latent Dirichlet Allocation) 모형 등이 있으며[10], 본 연구에서는 수집된 문서들 간의 상호의존성(inter-dependency)과 감사이슈의 포괄성(diversity)을 잘 반영하는 ‘LDA 모형’을 채택했다. LDA 모형의 기본구조와 말뭉치(corpus) 생성과정을 도식화하면 <Fig 3.>과 같다.



<Fig. 3> A basic structure and corpus generation process of LDA model

LDA 모형은 Blei et al.(2003)가 문서들의 주제를 분류하기 위해 최초로 고안하였는데[14], 발견되지 않은 그룹에 의해 설명될 수 있는 관측치(observation)의 집합을 설명하는 생성모델(generative model)이라 할 수 있다. LDA 모형은 잠재적 확률추정기법을 사용하며, 이 기법은 하나의 문서는 여러 개의 주제를 포함하거나 여러 문서가 공통의 주제를 공유할 수도 있다는 점을 전제로 한다. 각 주제는 일정한 분포를 가지고 있고, 각 주제의 분포는 데이터 전체 분포의 하위집합으로 구성되는 위계모형(hierarchical model)의 특성을 지닌다. LDA 모형의 기본적인 구조와 문서 생성과정은 <Fig. 3>과 같다.

LDA는 주어진 개수의 주제에 대하여, 어떤 특정한 분포에 따라 각 단어가 주제별로 출현할 빈도의 분포와 문서별로 각 주제를 포함하는 경우의 분포 간 곱을 각 주제에 할당된 단어의 빈도로 나눈 값을 기준으로, 해당 단어가 어떤 주제에 할당되는지를 깃스 샘플링(Gibbs sampling)[15]을 통해 반복적으로 계산하여 각 주제를 구성하는 단어들의 집합을 확률적으로 추정해내는 모형이다. 도식적으로 설명하자면, <Fig. 3>과 같이 문서(D), 단어 집합(N_d)이 주어졌을 때, 확률 벡터(ϕ_i)에 의해 특징이 결정되는 각 단어에 대한 분포로써 주제(t)를 정의할 수 있다. 각 문서는 주제에 대한 분포(θ_j)와 결합되어 설명된다. 주제에 대한 단어의 확률값 ϕ_i 는 파라미터 β 의 영향을 받고, 문서에 대한 주제의 확률값 θ_j 는 파라미터 α 에 의해 영향을 받는 은닉 변수(latent variable)이다.

모형이 깃스 샘플링을 통해 최적화되는 동안 현재의 확률 분포 상태에 대한 품질은 엔트로피를 활용한 혼잡도(perplexity)의 수치가 얼마나 빠른 시간 내에 효과적으로 안정화되는가를 기준으로 측정한다.

$$perplexity(w) = exp \left\{ \frac{1}{N} \sum_d \log \left(\sum_w \theta_j^d \phi_i^j \right) \right\}$$

본 연구에서는 OwlNest and Audit and Inspection Research Institute(2015)[12]에서 적용한 바와 같이,

혼잡도를 기준으로 모형을 측정했을 때 일반적으로 한국어 문서의 주제분석에서 빠르고 효과적인 최적화로 좋은 성능을 나타내는 것으로 판단되는 비대칭-대칭 혼합 어휘소 기반(Asymmetric-Symmetric Lexeme-based) LDA 모형을 사용했다. 이 모형은 비대칭 파라미터 α' 을 추가함으로써 문서-주제간의 분포 최적화에 이점이 많은 것으로 보고되고 있다[16].

3. 경계추정(Boundary estimation) 단계

1) 플랫폼별 감사대상 노출빈도(Exposure frequency of the audit object by platform) 분석

플랫폼별 감사대상 노출빈도는 ‘1. 포화표본 추출 단계’의 ‘분류체계’에서 설정한 ‘감사대상 관련 키워드’를 포함하는 보도 자료 또는 인터넷 상의 문서가 플랫폼별로 얼마만큼 빈도 높게 나타나는지를 산출한 값이다. 이 노출빈도는 감사이슈 판단 시 플랫폼에서 감사대상이 얼마나 높은 빈도로 노출되었는지를 참고할 수 있는 양적 지표로 사용될 수 있다. 따라서 그 양만큼 국민 또는 이해관계자에게 많이 알려져 있는 정책이슈라고 해석할 수 있다. 또한 감사대상 노출빈도뿐만 아니라 감사대상에 대한 총 노출빈도 대비 부정적 견해 노출빈도로 나타나는 감사대상에 대한 부정적 견해 노출비율도 측정했다.

본 연구에서는 전술한 플랫폼 분류에서와 같이 일반플랫폼은 중앙미디어와 지방미디어를 구분하여 감사대상 노출빈도를 측정했다. 중앙미디어 플랫폼에서는 ‘네이버’ 뉴스 포털의 노출빈도, ‘다음’ 뉴스 포털의 노출빈도를 측정하였으며, 지방미디어 플랫폼에서는 ‘미디어가온’ 지방지 포털의 노출빈도를 측정했다. 전문 플랫폼은 전문미디어와 전문저널을 합쳐서 노출빈도를 측정했다.

플랫폼별 노출빈도 분석결과를 종합하면 <Table 3>과 같다. 결과를 보면 ‘미디어가온’ 지방지 포털에서는 ‘네이버’, ‘다음’ 뉴스 포털보다 ITS에 대해 정책이슈를 많이 내포하고 있다. 특히, ‘2-① 교통정보 연계’는 ‘네이버’ 뉴스 포털, ‘다음’ 뉴스 포털 그리고 ‘미디어가온’ 지방지 포털에서 모두 전체

보도자료 중 부정적 견해를 포함하는 보도자료의 비율이 70% 이상으로 고위험 정책대상(이하, 고위험 정책대상)으로 분류됐다. 그리고 ‘1-⑤ 차세대 지능형 교통시스템(C-ITS)’은 ‘미디어가온’ 지방지 포털과 전문 플랫폼에서 고위험 정책대상으로 도출됐다.

(Table 3) A Summary on the exposure frequency of the audit items in each platform

Platforms Analysis Indicators	Public Platform			Professional Platform
	「Naver」 News Portal	「Daum」 News Portal	「Media Gaon」 Local News Portal	
Frequency of audit item exposure	142,740	258,734	7,625	937
Number of negative comments	79,951	126,042	5,682	418
Percentage of negative comments (%)	56.0	48.7	74.5	44.6
High-risk audit items*	1-①, 1-⑥, 1-⑦, 2-①, 2-②	1-③, 1-⑧, 2-①, 2-③	1-④, 1-⑤, 1-⑥, 1-⑦, 2-①, 2-⑤	1-⑤

* The high-risk audit items with over 70% of negative comments in general platform or those with over 50% of negative comments in the professional platform are drawn. The underlined shows audit items that are found to have high-risk for 3 times or more in the each platform.

1-1) 대중 플랫폼 노출빈도 분석결과

1-1-1) 「네이버」 뉴스 포털의 노출빈도 분석결과

2012년 1월부터 2014년 12월까지 「네이버」 뉴스 포털에 게시된 보도 자료 중 <Table 1>의 감사대상 키워드를 포함하는 보도 자료의 노출빈도는 142,740건으로 나타났다. 이 중 부정적 견해를 포함하는 보도 자료는 79,951건으로 56.0%의 비중을 차지하였다. 또한 고위험 정책대상은 ‘1-⑥ 교통사고분석 시스템(TAAS)(72.6%)’, ‘1-① CCTV 주정차 단속시스템(71.6%)’, ‘1-⑦ 자동요금징수시스템(ETCS) (70.3%)’ 등으로 분석됐다.

1-1-2) 「다음」 뉴스 포털의 노출빈도 분석결과

2012년 1월부터 2014년 12월까지 「다음」 뉴스 포털에 게시된 보도 자료 중 <Table 1>의 감사대상 키워드를 포함하는 258,734건이고, 이 중 부정적 견해를 포함하고 있는 보도 자료는 126,042건으로, 48.7%의 비중을 차지했다. 또한 고위험 정책대상은 ‘2-② 교통정보 연계(78.8%)’, ‘1-⑧ 교통안전정보관리시스템(TMACS)(78.6%)’, ‘2-③ 교통정보 중복(73.7%)’, ‘1-③ 디지털운행기록계(DTG)(73.5%)’ 등으로 분석됐다.

1-1-3) 「미디어가온」 지방지 포털의 노출빈도 분석결과

2012년 1월부터 2014년 12월까지 「미디어가온」 지방지 포털에 게시된 보도 자료 중 <Table 1>의 감사대상 키워드를 포함하는 보도 자료는 7,625건으로 이 중 부정적 견해를 띄는 보도 자료는 5,682건으로 74.5%의 비중을 차지하는 것으로 나타났다. 이는 「네이버」 뉴스포털 보다 18.5%, 「다음」 뉴스포털 보다 25.8% 높게 나타난 결과로, ‘지방미디어’ 플랫폼이 ‘중앙미디어’ 플랫폼에 비해 고발성 보도 자료의 빈도가 높음을 의미한다. 또한 고위험 정책대상은 ‘2-⑦ 자동요금징수시스템(ETCS)(88.9%)’, ‘2-① 교통정보 연계(80.0%)’, ‘1-① CCTV 주정차 단속 시스템(77.8%)’, ‘1-⑤ 차세대 지능형 교통시스템(C-ITS)(76.3%)’, ‘1-⑥ 교통사고분석시스템(TAAS)(75.0%)’, ‘2-⑤ T-map 등 교통 앱 활용(75.0%)’, ‘1-④ 버스안내정보시스템(BIS)(73.7%)’ 등으로 분석됐다.

1-2) 전문 플랫폼 노출빈도 분석결과

2012년 1월부터 2014년 12월까지 전문미디어에 게시된 자료 중 ITS관련한 자료는 937건으로 이중 부정적 견해를 띄는 자료는 418건으로 44.6%의 비중을 차지하고 있었다. 이는 「네이버」 포털, 「다음」 포털, 「미디어가온」 지방지 포털 보다 낮은 수치로, 전문 플랫폼에서는 정책대상에 대한 부정적 견해를 띤 보도자료 비중이 낮게 나타남을 의미한다. 그리고 부정적 견해를 띤 보도자료 비중이 70% 이상인 고위험 정책대상은 존재하지 않았으나 50%이상

70%이하에 해당하는 정책대상은 ‘1-⑤ 차세대 지능형 교통시스템(C-ITS)(52.3%)’으로 나타났다.

2) 감사이슈 판단

텍스트마이닝을 이용한 정책이슈 탐색기법의 마지막 단계인 ‘감사이슈 판단’ 과정에서는 LDA 모형을 이용한 주제분석 결과 도출되는 여러 가지 분석지표를 활용하여 정책적으로 고려할 필요가 있는 이슈들을 판단한다. 본 연구에서는 전술한 바와 같이 감사대상의 노출빈도와 주제분석 결과 도출되는 주제어와 문서군별 주제비중 그리고 문서별 주제비중 등을 판단지표로 활용했다.

앞의 플랫폼별 노출빈도 분석에서 「네이버」 뉴스 포털, 「다음」 뉴스 포털 그리고 「미디어가온」 지방지 포털에서 모두 고위험 정책대상으로 분류된 ‘2-① 교통정보 연계’의 경우를 예시로 감사이슈를 판단하는 방법을 제시하면 다음과 같다.

본 연구에서는 주제분석을 통해 도출된 결과를 <Attachment 2>와 같은 형태로 표출했다. 맨 왼쪽 열에는 주제 비중 값으로 정렬한 순위 값, 두 번째 열에는 문서군별 주제비중(topic weights by document cluster)이 나타나게 했다. 그리고 3~6열에는 문서군 중에 해당 문서군에 기여도가 가장 높은(=문서 주제비중(topic weight by each document)이 가장 높은) 대표 보도자료의 매체명(3열), 보도자료 제목(4열), 게시일자(5열), URL(6열)을 표출하도록 했다. 그리고 마지막 7열에는 문서군의 주제어(topic words)를 제시하도록 했다. 이러한 양식화된 주제분석 결과 테이블을 통해 분석가는 해당 정책의 이슈를 종합적으로 파악할 수 있다. 주제 비중을 통해 문서군이 대표하는 주제의 가중치를 양적으로 판단할 수 있으며, 주요 정책이슈에 대한 판단은 대표 보도자료, 문서군의 주제어를 통해 정성적으로 판단할 수 있으며, 또는 주제비중이 높은 문서의 내용분석을 통해서도 파악할 수 있다.

3) 감사이슈 탐색 결과

감사이슈 탐색 결과 <Attachment 3>과 같다.

‘개발 및 구축’ 분야의 경우 1-① ‘CCTV 주정차 단속시스템’의 획일적 도입으로 인한 상권 위축과 지역의 특성을 반영한 탄력적 운영을 요구하는 등 9건의 정책(감사)이슈가 주요이슈로 선정됐다.

이뿐 아니라 ‘운영 및 관리’ 분야의 경우 2-① ‘교통정보 연계’이슈로 경찰의 UTIS 사업과 국토부 ATMS 사업과의 중복투자문제 등 5건의 정책(감사)이슈가 주요이슈로 선정됐다.

V. 결론

본 연구는 ‘빅데이터’로 구축된 플랫폼(인터넷, 전문DB 등)을 정책 수립 및 정책형성의 중요한 정책 수단으로 활용하고자, 경계분석이론과 텍스트마이닝(Text Mining)을 이용하여 국내 인터넷환경에 널리 구축된 비정형적인 빅데이터로부터 정책이슈를 탐색하는 기법을 정립했다.

또한, 정립한 탐색기법을 이용하여 도로부문 지능형교통체계(ITS)를 대상으로 사례적용을 수행함으로써 그 활용도를 테스트한 결과, ‘네이버’ 뉴스, ‘다음’ 뉴스, 전문 문헌 등 국내 인터넷에 공개된 자료 등을 이용하여 한정된 자원과 시간 내에 빠르게 정책(감사)이슈를 발굴(data-driven issue tracking)할 수 있었다.

아울러, 중앙미디어(2곳), 지방미디어(1곳), 전문미디어(1곳)에 대한 탐색결과, 부적정 견해를 비중이 70%이상으로 고위험 정책(감사)이슈로 구분할 수 있는 고위험 대상을 탐색하였으며, 앞으로 이러한 이슈들에 대한 사후관리도 중요하다고 판단된다. 연구의 한계 및 향후 연구과제로는 본 연구의 분석기법이 인터넷(웹)을 통해 가용할 수 있는 미디어 보도 자료를 대상으로 하는 탐색적 접근(descriptive approach)방법에 주로 의존하였으나, 향후에는 전문가 의견을 보완하는 경험적 접근(heuristic approach) 방법도 함께 고려할 필요가 있을 것으로 판단된다.

REFERENCES

- [1] Oh C. S.(2016), *A Study on the Audit Issues Investigation Method Applied Bigdata and Textmining - Focusing on the Project to Construct Transportation Information in the Road sector*, Audit and Inspection Research Institute.
- [2] OECD(2015), *Government at a Glance 2015*, Organization Economic Cooperative Development Publishing, p.146.
- [3] GAO(2015), *Open Data for Government Accountability, Efficiency, and Effectiveness*, GAO-15-240CG, Government Executive Media Group of the United States Government Accountability Office.
- [4] ECOPA(2014), *Risk-Based Audits: Principles and Implementation, Optimal Audit Case Selection*, White Paper, The European Consensus Platform.
- [5] Dunn W.(2008), *Public Policy Analysis*, 4th edition, Pearson, p.97.
- [6] Dunn W.(2012), *Public Policy Analysis*, 5th edition, Pearson, p.90.
- [7] Hearst M.(1999), *Untangling Text Data Mining*, Retrieved March 20, 2007, from School of Information, University of California, Berkeley.
- [8] Hearst M.(2003), *What is Text Mining?*, Retrieved March 20, 2007, from School of Information, University of California, Berkeley.
- [9] Grabowski M. and Lula P.(2011), “Automatic Identification of Time and Subject Related Patterns in Large Collection of Scientific Publications,” *Management Information Systems*, vol. 6, no. 3, p.1.
- [10] Jung D. M.(2015), *Study of Topic Classification and Keywords for Online Fiction Search*, A Master’s Thesis of Seoul National University, p.22.
- [11] Kim D. G. and Weon J.(2013), *Development of Foresight Method for Future Disaster Through the Analysis of Complex Foresight*

Methodology, National Disaster Management, pp.59-79.

[12] OwlNest and Audit and Inspection Research Institute(2015), *Audit Topic Model for Transportation Information System using Textmining*, p.42.

[13] Jung D. M., Kim J. S., Kim K. N., Heo J. W., Ohn B. W. and Kang M. J.(2013), "A Proposal of a Keyword Extraction System for Detecting Social Issues," *J. Intell. Inform. Syst.*, vol. 19, no. 3, pp.1-23.

[14] Blei D. M., Ng A. Y. and Jordan M. I. (2003), "Latent Dirichlet Allocation," *Journal of Machine Learning Research* 3.

[15] Steyvers M. and Griffiths T.(2007), *Probabilistic Topic Models*, In Landauer, T., McNamara, D. S., Dennis, S. and Kintsch, W. (Eds.), *Handbook of Latent Semantic Analysis*, Hillsdale.

[16] Wallach H., Mimno D. and McCallum A.(2009), *Rethinking LDA: Why Priors matter*, NIPS.

<Attachment 1> Negative meaning dictions adopted in this study(total 270 dictions)

가공, 가로채, 가중, 가짜, 간권, 갈등, 감싸실, 감싸원, 감을관계, 거꾸로, 검은돈, 겪는, 겪다, 겪은, 경제적이지 못한, 계층 이기주의, 고무줄, 고압적, 고양이, 고의, 고장, 공간, 공공연한, 공공연한 비밀, 공포, 과다/과소계상, 과잉감사, 관행, 구색 맞추기, 권위적, 권익, 그것이 알고 싶다, 그들만의 리그, 그들만의 잔치, 그럴리가, 근본원인, 금지초문, 꼬리 자르기, 꿀꺽, 나눠먹기, 나빠요, 나쁘다, 낙하산, 날릴 판, 남용, 납세자, 낭비, 널뛰기, 논란, 누락, 눈가리고 아웅, 눈뜯장님, 눈먼 돈, 능장, 능장행정, 능률적이지, 대오각성, 덮쳐, 도덕불감증, 도덕적 해이, 도산, 돌아가고, 둔감, 뒷북, 드러났다, 맹물, 맹탕, 먹통, 먼지털이식, 멋대로, 면죄부, 목적 없는, 목적의, 못들은 체, 무리한, 무사안일, 무소불위, 무용지물, 묵묵부답, 묵인, 문란, 문제, 문제점, 묻지마, 미래세대, 미미, 미실시, 미흡, 민심불감증, 민원, 밀빠진 독, 박탈, 방만, 복마전, 복지부동, 부당, 부실, 부의부빈익빈, 부자위, 부재, 부적절, 부적정, 부정확, 부족, 부진, 부처이기주의, 부패불감증, 부풀려, 부풀리기, 불감증, 불만, 불법, 불편, 불편한 진실, 비경제, 비능률, 비리, 비밀, 비효과, 비효율, 빙산의, 뺨질관 국, 뺨튀기, 사각, 사기저하, 사실무근, 사실상, 사후약방문, 생선, 선무당, 선의, 설거지, 설익은, 성역, 소리없는 아우성, 소외, 소외고 외양간 고치기, 소홀, 속빈강정, 숨방망이, 수법, 쉬쉬, 시급, 시급하다, 시급한, 시급할, 식물, 신뢰성, 신중함, 신중한 입장, 실패, 심각, 심각한가?, 쌓기, 쓴데없는, 안전불감증, 알려졌다, 알려진 비밀, 압력, 약자, 어디서 뭐하고, 어이없는, 어찌라고, 억울, 엉뚱한, 엉망, 엉터리, 여전, 엄불, 예견된, 예고된, 예산낭비, 오락가락, 오용, 오작동, 예곡, 외양간 고치기, 우려, 울고, 울다, 울상, 위반, 위배, 위헌, 유명무실, 윤리, 은혜, 의혹, 이기주의, 이렇 수가, 이제 와서, 이중, 이중잣대, 일관성, 있으나마나, 재발, 잣말, 저인망식, 저조, 저축, 저하, 적발, 적시성, 전략, 절실, 점검, 정치, 정확성, 제각각, 제식구 감싸기, 조작, 졸속, 주먹구구, 주인 없는, 줄줄, 중복, 지금까지, 지연, 지적, 지지, 지지부진, 진짜?, 집단이기주의, 징계, 짜고치는, 짜고치는고스톱, 찜짜미, 차질, 착복, 찡찡다, 찡기다, 쟁전, 처벌, 철저한, 청부감사, 청탁, 축소, 침묵하는 아우성, 칸막이, 코드, 큰문제, 탁상, 태만, 털어는, 털어냈다, 텅텅, 토로, 특혜, 파문, 판국, 평평, 편법, 편취, 표적, 필요하다, 필요한, 하나하나, 해바라기, 허덕, 허위, 혈세, 형식적, 형평성 없는, 형평적이지 않은, 혼란, 확인되었다, 환경, 횡령, 효과적이지 않은, 효율적이지 않은, 훼손

<Attachment 2> A Example of topic analysis(in the case of '2-① Traffic Information Connections')

순위	주제비중	대표 보도자료			문서군의 주제어
		매체명	보도자료 제목	URL	
1	57.14	서울경제	[전국 페트롤] 서울버스·지하철 실시간 위치정보 웹 포털 공개	http://news.naver.com/main/read.nhn?mode=LSD&mid=sec&sid1=102&cid=011&aid=0002452961	코스, 구간, 개통, 하천, 인양천, 불법, 산림, 울재길, 구로구, 분야, 부시장, 명예, 인근, 길, 목감천, 도형진, 만화, 강남구, 성남기, 개
2	50.00	교통신문	'버스·지하철 운행정보 한눈에'	http://www.gyotongan.com/news/articleView.html?ikno=59713	정보, 영상, 국토, 부, 있다, 실시간, 나쁜(게이시, 교통, 사이트, 기법, 국토, 웹, 스마트폰, 상황, 콘텐츠, 신규, 불법, 개발자, 일부, 교통사, 지
3	32.60	교통신문	내년부터 서울 지역 주차장 찾기 쉬워진다	http://www.gyotongan.com/news/articleView.html?ikno=49340	공영, 전광판, 실시간, 제공, 여부, 인프라, 주차, 부실, 내달, 통합, 찾다, 지역, 서울
4	17.67	정책브리핑	이론자들과 경사 시행... 공공임대주택 도입	http://media.daum.net/society/newsview?newsid=2013122716400699	건축물, 경우, 지원, 주택, 대상, 완화, 건축, 확대, 행위, 이하, 휴대, 사무소, 모기지, 탄발, 운수, 수익, 규제, 위헌, 선택, 오건
5	17.39	교통신문	서울버스운행정보(태터)와 활용민간업체 확대된다	http://www.gyotongan.com/news/articleView.html?ikno=50245	공영, 전광판, 실시간, 제공, 여부, 인프라, 주차, 부실, 내달, 통합, 찾다, 지역, 서울
6	16.70	연합뉴스	<생활도 잡자> 해결 최첨단 지능형 해상교통시스템 사용 때 '덕분'	http://news.naver.com/main/read.nhn?mode=LSD&mid=sec&sid1=102&cid=001&aid=0006891760	해결, 사고, 장비, 신발, 상황, 모나터링, 정보, 분석, 상황, 지능, 상황실, 첨단, 알리다, 위급, 운항, 해상수상부, 도울, 풍스, 지, 확인
7	4.80	뉴스시스	LX공사, 무인항공기 활용 공간 정보산업 활성화 포럼	http://media.daum.net/society/newsview?newsid=20141128164509469	판장, 유람선, 수상물, 밀수, 운항, 어시장, 투입, 현대, 조선, 관공래, 도개, 추진, 해양, 공간, 대형, 상정, 전망대, 데크, 관공래, 스, 장비
8	3.20	정책브리핑	각 부처 서비스(서비스)연계·협조체계 구축	http://media.daum.net/society/newsview?newsid=20120307121005234	크리드, 전략, 함진, 참여, 실증, 설비, 정전, 배전, 지능, 운영, 인프라, 발전, 선택, 거점, 단속, 구간, 전기, 획득, 인출, 용
9	2.56	디지털타임스	서울시, 5월 버스정류장에서 지하철 운행정보 제공	http://news.naver.com/main/read.nhn?mode=LSD&mid=sec&sid1=105&cid=029&aid=0002115527	외교, 서울, 문화, 현장, 물가, 공정위, 의학, 교육부, 하다, 특, 필통, 왕십리, 입법, 연설회, 국회, 여수, 건강, 연구원, 기업, 공개
10	2.40	정책브리핑	[서울신문] 지류스, 여러차례 할로이탈 감사기법은 없어	http://media.daum.net/politics/newsview?newsid=20140505181005865	판장, 유람선, 수상물, 밀수, 운항, 어시장, 투입, 현대, 조선, 관공래, 도개, 추진, 해양, 공간, 대형, 상정, 전망대, 데크, 관공래, 스, 장비

〈Attachment 3〉 The major issues investigated by sector after the result of topic analysis

먼저 도로부문 ITS 구축사업 ‘개발 및 구축’ 분야의 주요 감사이슈는 다음과 같다.

1-① ‘CCTV 주정차 단속시스템’과 관련해서는 지방자치단체의 버스 탑재형 이동식 주정차 단속시스템 도입이 주요 이슈로 떠오르는 가운데, 시스템의 획일적 도입으로 인한 상권 위축 등의 문제와 지역의 특성을 반영한 탄력적 운영을 요구하는 의견들이 제기됐다. 1-② ‘첨단교통관리시스템(ATMS)’과 관련해서는 ATMS와 통행료자동요금징수시스템(ETCS)은 대중교통서비스인 버스정보시스템(BIS)이나 대중교통요금 전자지불시스템(EFCS)에 비해 글로벌 경쟁력이 비교열위에 있다는 문제점이 비중에 있게 지적됐다. 1-③ ‘디지털 운행기록계(DTG)’과 관련해서는 DTG 운행기록 조작, DTG 성능검증 강화 필요, 화물차 기기 장착률 저조, 터널 등 특수구간에서의 미작동 등의 문제가 지적됐다. 1-④ ‘버스정보안내시스템(BIS)’과 관련해서는 지하철 등 타 교통수단과의 연동 문제 제기됐다. 1-⑤ ‘차세대 지능형 교통시스템(C-ITS)’과 관련해서는 국토교통부의 20조 원 투자계획 등을 앞두고 국외 대비 우리나라 기술 관련 법·규제 등을 점검할 필요가 있다고 언급됐다. 1-⑥ ‘도로교통공단 교통사고분석시스템(TAAS)’과 관련해서는 사고 발생 시 경찰신고의 과소보고 문제와 교통사고 다발지역을 중심으로 도로시설물 파손 행위가 증가하고 있으나 원인자 적발의 현실적 어려움으로 지방자치단체 예산에 부담 가중 문제가 지적됐다. 1-⑦ ‘자동요금징수시스템(ETCS)’과 관련해서는 대기업의 교통카드 사업 해외진출에 따라 국제규격 표준화 준수 문제가 언급됐다. 1-⑧ ‘교통안전정보관리시스템(TMACS)’과 관련해서는 동 시스템 관련 기업의 해외진출 소식이 많이 언급되고 있었으나, 특이한 감사이슈는 발견되지 않았다. 1-⑨ ‘도시교통정보시스템(UTIS)’과 관련해서는 UTIS 정보 수집률에 대한 신뢰성과 일부 지방자치단체에서 구축 장비 오작동이 빈도 높게 언급됐다.

다음으로 도로부문 ITS 구축사업 ‘운영 및 관리’ 분야의 감사이슈는 다음과 같다.

2-① ‘교통정보 연계’와 관련해서는 세월호 침몰 과정에서 VMS를 전혀 활용하지 못했으며, 경찰의 UTIS 사업과 국토부 ATMS 사업과의 중복투자가 문제점으로 언급됐다. 2-② ‘교통관리센터’와 관련해서는 운영상의 문제점으로 도시교통정보센터(UTIS) 기능 상실과 인천 등 일부 지방자치단체에서의 콜택시와 연계된 UTIS 오작동 문제가 제기됐다. 2-③ ‘교통정보 중복’과 관련해서는 공공데이터 개방정책에 힘입어 공공기관 앱이 우후죽순으로 늘고 있으나 유사한 서비스 기능이 있는 민간 앱과 40%나 중복된다는 문제점이 지적됐다. 2-④ ‘교통정보 활용’과 관련해서는 국가에서 수집·제공하는 교통정보와 민간에서 수집·제공하는 교통정보를 융합한 민·관 융합 교통정보가 2015년 말경 등장할 것이나, 이에 따른 교통정보의 정확도 및 신뢰성 검증의 문제가 남아 있다고 언급됐다. 2-⑤ ‘교통정보 앱 활용’과 관련해서는 공공기관 정보를 활용한 생활형 교통정보 앱이 많이 출시되고 있으나 버스 노선 변경 등 중요 생활형 정보 미제공에 따른 시민 불편 등의 문제와 관련 정보를 제공하는 UTIS의 기능 상실로 인해 관련 앱과의 연동성에 문제가 있다고 지적됐다.

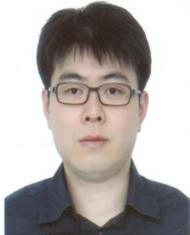
저자소개



오 창 석(Oh, Chang-Seok)
2015년 5월~현재 : 감사원 감사연구원 행정사무관(연구관)
2014년 8월 : 서울대학교 건설환경공학부 박사(교통공학전공)
2005년 9월~2015년 5월 : 감사원 감사연구원 행정주사(연구관보)
2004년 2월~2005년 9월 : 한국도로공사 도로교통연구원 연구원
e-mail : brandonoh@korea.kr



이 용 택(Lee, Yong-taek)
2016년 1월~현재 : 감사원 행정안전감사국 수석감사관
2005년 4월~2015년 12월 : 감사원 국토환경감사국, 지방행정감사국 부감사관 등
2004년 2월~2005년 3월 : 서울지방경찰청 교통개선기획실장
2003년 8월 : 서울대학교 대학원 환경계획학과 박사(교통공학전공)
e-mail : icarus01@korea.kr



고 민 수(Ko, Minsu)
2016년 2월 : 한국과학기술원 박사(전산학, 수료)
2011년 12월 : 지식경제부 Software Maestro
2010년 8월 : 서울대학교 석사(전산언어학)
2006년 2월 : 서강대학교 학사(독어학)
e-mail : ryan0802@kaist.ac.kr