pISSN : 2508-3384 eISSN : 2508-3392
https://doi.org/10.22640/lxsiri.2018.48.1.15
Journal of Cadastre & Land InformatiX Vol.48 No.1 (2018)
한국국토정보공사「지적과 국토정보」제48권 제1호 2018년 6월 p. 15-33

# A Study on Identifying Topics and Trends in International Cadastral Research Using LDA: With Special Reference to the FIG Peer Review Journal

## LDA를 이용한 국제지적연구의 주제와 추세확인에 관한 연구: 특히 FIG Peer Review Journal을 중심으로

kim, Yun-Ki*
김 윤 기

**Abstract**

The main purpose of this study was to identify the topics and research trends of international cadastral research using LDA. To achieve this goal, I reviewed the literature on LDA and international cadastral study and formulated four research questions that are topics of cadastral researchers, distribution of topics, the most influential topics and changes of topics over time. To answer these research questions, I analyzed 370 papers published in the FIG Peer Review Journal between January 1, 2008, and October 31, 2017, using LDA. As a result of the analysis, I confirmed that there are twelve major topics in international cadastral research. And the most influential topic of these topics was identified as topic 2(cadastral information systems), and topic 5(land development and land administration) was also confirmed as playing an important role in the overall document. These two topics have been the most popular topics whose trendlines have been very active over the past decade and will play a leading role in future cadastral research.

Keywords: Topics, Trends, LDA, International Cadastral Research, Cadastral Information Systems, Land Development, Land Administration

## 1. Introduction

We cannot live without land. In other words, humans have to solve all the food, clothing and shelter problems on land. We eat food grown on farms, make clothes using raw materials from plants grown on the ground, and build homes on plots of land. The land has always played a key role in supporting human life throughout history. The land was not used densely in the agricultural

* 청주대학교 지적학과 교수 Professor, Department of Land Management, Choengju University (kim2875@cju.ac.kr)

age. In this era, the concepts of land boundaries were less evident than now, so farmers often encroached on their neighbors' land without knowing it. However, the boundaries between neighboring parcels became clearer as industrial society began to emerge. In this age, the land was used more densely than agricultural age, so land prices skyrocketed, and land boundary disputes frequently occurred in neighboring areas. Cadastral research in this period was mainly focused on solving boundary disputes and managing land information efficiently. As the IT revolution spread around the globe in the 2000s, many researchers concentrated on the technological aspects of cadastre. Notably, cadastral information systems have attracted many scholars' attention. Besides, many researchers have tried to suggest ways to solve various problems that have arisen in the land development processes. Also, while some scholars have studied the legal aspect of cadastre, other researchers have focused on cadastral education. As the three-dimensional use of land became more and more common, scholars showed a keen interest in 3D cadastre research(Kim et al., 2006; Paulsson et al., 2011).

The topics of cadastral research have been continuously changed to meet the demands of the times. However, there were not many scholars who studied the topics of cadastral research and trends intensively. Only a few researchers have used mere descriptive analysis or content analysis to identify topics and research trends in the field of cadastral research(Karpik et al., 2016). Recently, big data analysis techniques have been rapidly developed. Thanks to these new analytical techniques, researchers can analyze data at speeds that were unimaginable in the past. Especially, LDA allows researchers to detect topics and analyze trends in large documents(Blei et al., 2003). For this reason, scholars from diverse fields today use LDA to identify research topics and trends in their areas(Griffiths et al., 2004). LDA can also be a great help in detecting research topics in cadastral research and identifying research trends. In particular, it is possible to obtain per-word topic allocations, per-document topic proportions, and per-corpus topic distributions in cadastral studies using LDA. In this study, I try to identify the topics and trends of international cadastral research using LDA.

## 2. Literature Review

### 2.1. LDA

#### 2.1.1. LDA Process

LDA is a generative probability model of a corpus(Blei et al., 2003). LDA assumes that a word has a meaning in relation to other words. People tend to use similar terms when talking about the same topic, so similar terms tend to co-occur in documents on the same topic. LDA explains per-word topic assignment, per-document topic proportions, and per-corpus topic distributions using the probabilities of word co-occurrences. LDA can estimate the topics of new documents without updating the current model. LDA is an algorithm that finds hidden variables based on observed variables(Hoffman et al., 2010). In general, the hidden variables model has a structured

distribution in which observed data interacts with random variables(Lafferty, 2006). We can find the interaction between these observed variables and the hidden topic structure in the probabilistic generation process related to LDA.

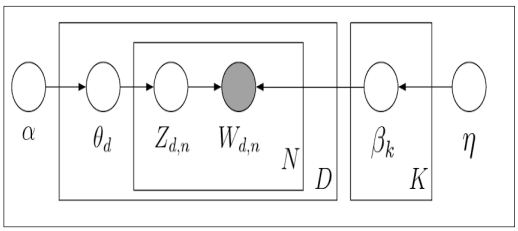


Figure 6. LDA Process

Source: Blei, 2003

Figure 1 shows the LDA process as a graphical model(Mcauliffe et al., 2008). Boxes are plates that represent replicates. The outer box represents documents and the inner box describes the repeated selection of topics and words within a single document(Blei et al., 2003). The hidden topic structure in a corpus is displayed as hidden random variables. In this figure, nodes represent random variables, while edges represent dependence between random variables, respectively(Zhu, 2014). Shaded nodes also represent observed random variables and unshaded nodes represent hidden random variables(Blei et al., 2003). Proportions parameter is denoted as $\alpha$ and topic parameter is represented as $\eta$. Topics are represented as $\beta_k$ and topic proportions per document is denoted as $\theta_d$. Topic assignments per word is indicated by $Z_{d,n}$ and observed word is represented as $W_{d,n}$ in this model(Chang et al., 2009). As can be seen in this figure, LDA variables consist of three levels(Mcauliffe et al., 2008). Its first level is corpus-level, where parameter $\alpha$ and parameter $\eta$ are assumed to be sampled once during the process of corpus generation and $\theta_d$ is a document-level variable that is sampled once for each document. Finally, $Z_{d,n}$ and $W_{d,n}$ are word-level variables, which are sampled once for each word in each document(Blei et al., 2009).

And the direction of edges in this figure means dependence. That is, the topic proportions per document depend on proportions parameter, the topic assignment per word depends on the topic proportions per document, and observed word depends on the topic assignments per word. Also, the topics are affected by topic parameter, and observed word is affected by the topics. As can be seen above, the observed word is affected by all the nodes in this figure. In other words, this is a kind of dependent variable that is affected by all the variables in the model. In this model, we know only proportions parameter, topic parameter, and the observed word. We need to infer hidden variables such as topic distributions per corpus, the topic proportions per document, and the topic assignments per word with these known variables. This process is called LDA inference and collapsed Gibbs sampling is often used for this. LDA is a kind of a mixed membership model. This method differs significantly from the existing mixed membership models, which restrict a single topic to a single document. In other words, this model assumes that unlike the existing model, a document can have more than one topic.

2.1.2. Literature Review

Even before topic modeling was used, researchers had used a variety of methods to analyze papers from journals. However, existing analytical methods have not been very helpful in finding research topics and research trends in academic fields. To overcome the drawbacks of these existing methods, researchers tried to use LDA. Many scholars analyzed the abstracts of articles to discover research topics and research trends in their fields. For example, Griffiths et al.(2004) investigated the abstracts of papers in the Proceedings of the National Academy of Science(PNAS) from 1991 to 2001 using LDA. In particular, they compared the results of the analysis using LDA with those of the analysis using existing methods. The results reveal that their approach is very useful in identifying scientific topics. Blei et al. (2006) conducted topic detection on papers from Science from 1880 to 2000. They used LDA-based dynamic topic modeling to identify trends in research topics. Their model has proven useful in analyzing large unstructured data. LDA was also used to identify relationships between topics in medical research. In particular, Wu et al.(2012) utilized LDA to identify gene-drug relationships in biomedical research. To do this, they extracted and analyzed abstracts from MEDLINE. The results of their research show that the methods they used are more effective in identifying relationships among topics in the medical field than other methods. Gatti et al.(2015) analyzed the abstracts of papers published since the early 1950s in the field of operation research and management science. They extracted abstracts of 80,757 papers from 37 related journals and analyzed them using LDA. Their study results reveal that generality and speciality exists at the same time in each journal. Das et al.(2016) analyzed the papers published at the largest Transportation Research Board (TRB) Annual Meetings in the world using LDA. They used LDA to detect interesting research topics and research trends in transportation research. Sun et al.(2017) attempted to find out trends and topics of research in transportation using LDA. They tried to classify the topics of 17,163 papers in 22 transportation-related journals, published between 1990 and 2015, using LDA. The findings of their study show that there are 50 sub-topics in transportation research. Their findings also reveal that LDA is very useful in identifying the topics and trends in transportation research.

## 2.2. Identifying Topics and Trends in Cadastral Research

### 2.2.1. Topics in International Cadastral Research

There are various topics in international cadastral research. Recently, scholars have become increasingly interested in studying the technical aspects of cadastre(Paulsson et al., 2011). They have focused on cadastral information systems, geographic information systems, and land information systems. In fact, these topic are also frequently covered in the FIG Peer Review Journal. In addition, utilization of GPS is an important sub-topic in the technical aspects of cadastral research. Some scholars have tested the accuracy of the RTK service and

examined the possibility of applying it to cadastral surveying(Bisnath et al., 2013; Gučević et al., 2014). On the other hand, other researchers have focused on improving the cadastral infrastructure of RTK GPS(Rizos, 2007; Janssen et al., 2011).

Education is also a field of great interest to cadastral researchers in recent years. Today, cadastral systems and technologies are changing very rapidly day by day. For this reason, there is always a big demand for education in the cadastral field. Cadastral education is carried out in various forms not only in universities with cadastral departments but also in cadastral organizations. Such education is being actively carried out not only offline but also online. Using this social atmosphere, many researchers have tried to propose effective cadastral education models. In particular, Manzano-Agugliaro et al.(2016) compared the effectiveness of offline education and online education. Their findings show that the latter is more effective than the former. They proposed a new cadastral education model based on their research results.

In addition, researchers have focused on land administration systems. The land administration system takes different forms in different countries. In general, the land administration system of a country reflects its historical background. For this reason, scholars have attempted to study the land administration system of the country in which they live or to compare the land administration systems of various countries.

In particular, Steudler et al.(2004) suggested a basic framework for measuring and comparing the performance of land administration systems, and Silva et al.(2006) studied institutional barriers to implementing a land administration system in a developing country. Land governance and stakeholders are also areas of great interest to researchers in recent years. Especially, as the land use patterns of today become increasingly complex and diverse, many stakeholders have appeared in the land use process. A new form of land governance is needed to minimize conflicts among stakeholders and rationally implement land policies(Franco, 2010). In particular, Deininger et al.(2012) presented a framework for assessing land governance to identify good practices in the land sector.

Recently, many researchers have focused on natural disasters and sustainable cadastre. Today, there are huge natural disasters like earthquakes, hurricanes, and typhoons all over the world. When such a natural disaster occurs, existing cadastral information becomes almost meaningless. Thus, many scholars have tried to study sustainable cadastre to prepare for these disasters in advance. Enemark(2009) studied land administration infrastructures for the prevention and management of natural disasters. He confirmed that a sustainable land administration system plays a vital role in preventing natural disasters. 3D cadastre is a field on which researchers are focusing their attention. Researchers are interested in registering and managing underground space rights and air rights. In particular, Stoter et al.(2013) presented a new model for three-dimensional rights registration in the Netherlands.

### 2.2.2. Review of Related Literature

A few scholars have tried the topic identification

in the field of cadastral research. They attempted to detect research topics by analyzing the papers presented at conferences by using a simple method. For example, Paulson et al.(2011) studied the topic distribution of 3D cadastre research. They collected 105 papers on 3D cadastre published over the past ten years and analyzed it intensively. The results of their study show that there are four sub-topics in the 3D cadastre research: research on legal aspects, research on technical issues, research on registration, and research on the organization. Their findings reveal that among the sub-topics of 3D cadastre research, research on registration is the most important. However, the methods they used have limitations in that they can not identify topics accurately because they only use content analysis or descriptive analysis.

Oosterom(2013) categorized the topics of the papers presented at the second and third cadastre workshops and presented the direction of research in the future. The results of the research show that the topics that will lead 3D cadastre research in the future will be standardization, the implementation of full 3D cadastre, legal framework, the creation of 3D spatial unit, and the visualization of 3D cadastre. However, the study also has some limitations in the method of analysis. In other words, his paper only used a descriptive method in analyzing the papers on 3D cadastre presented at the workshops. Such a simple analysis alone will not be able to identify the topics of cadastre research accurately.

Karpik et al.(2016) attempted to analyze the research and actual trends of geospatial science using papers presented at the academic conferences of FIG(International Federation of Surveyors), ICA( International Cartographic Association), IPRS(International Society for Photogrammetry and Remote Sensing). Based on the results of the analysis, they found the most important topic in the academic activities of each institution. Their findings show that cadastre is the most essential topic in the academic activities of FIG, and generalization of spatial data is the most influential topic at ICA academic conferences. In ISPR's academic activities, scholars have shown the greatest interest in remote sensing. Their study can be highly appreciated in that they analyzed the papers published in the academic activities of the three cadastre-related institutions and detect topics and compare the results among the institutions. However, this study also has limitations in that it used the simple analysis method such as descriptive analysis.

### 2.2.3. LDA and Identifying Topics in Cadastral Research

There are also multiple topics in the field of cadastral research. Topics are rapidly changing as the research environment changes. The existing studies mainly used simple descriptive analysis or content analysis to detect research topics and research trends in the field of cadastre. However, it is difficult to precisely detect various topics in the entire cadastral research with such simple analysis techniques. In recent years, a vast number of academic papers have been published in the field of cadastral research. To process such large

amounts of data efficiently, researchers should use big data analysis methods such as LDA. Using LDA allows researchers to detect the topics in the corpus efficiently. With LDA, not only can we identify topics that exist throughout the corpus, but we can also detect topics that exist in each document in the corpus.

## 3. Research Questions

The main purpose of this study is to identify the topics and trends of international cadastral research using LDA. To lay the foundations for achieving this research objective, I conducted a literature review of LDA and international cadastral research. Based on these literature review results, I presented the following research questions.

Research Question 1: what are the topics that cadastral researchers are interested in?

In today's global cadastral research field, various topics are being studied. To understand the characteristics of a discipline properly, we need to know precisely what scholars in that area are mainly interested in. The field of global cadastral research is not much different from this. To detect topics of interest, we first need to identify per-corpus topic distributions and per-document topic proportions. Then we can use per-word topic assignments to identify specific topics for each word. Using this method, we can easily see what topics cadastral researchers are interested in.

Research Question 2: what are the topic proportions per corpus?

This research question relates to the distribution of topics throughout the corpus. In other words, researchers can use per-corpus topic distributions to determine which of these topics are influential. There are famous research topics in any discipline. In general, scholars are interested in these popular research topics and are studying them to keep up with others. We can often find this phenomenon in the field of cadastre research.

Research Question 3: what are the dominant cadastral research topics of each year?

Academic research is strongly influenced by our environmental changes. Cadastral studies that are closely related to our lives are affected more by environmental changes. The topics of cadastral research are also changing every year. We should use the per-document topic proportions to identify the research topics that are changing each year.

Research Question 4: how do cadastral research topics change over time?

To understand the direction of development of a research field, we need to grasp topics that are currently popular, but what is more important is to detect the trends of changes in the research topics. Favorite topics in the field of cadastral research are also changing every year. Once we understand these trends, we will be able to predict the future direction of the field of cadastral research.

## 4. Methodology

I visited the FIG website from November 1, 2017, to November 10, 2017, to collect papers on international cadastral research. I identified a total of 370 articles in the FIG Peer Review Journal using the resources tab on the FIG website. These

Table 1. Per-word topic assignments

| Topic 1 | Topic 2 | Topic 3 | Topic 4 | Topic 5 | Topic 6 |
|---|---|---|---|---|---|
| professional | data | energy | investment | land | nigeria |
| employers | system | marine | module | development | market |
| membership | paper | area | clearly | new | study |
| bodies | information | landuse | cadastral | using | networks |
| built | study | rail | cadastre | administration | protection |
| australia | model | israel | domain | rights | active |
| employees | based | test | office | building | loads |
| body | spatial | conversion | datasets | systems | classification |
| education | used | measurement | board | construction | reduced |
| students | property | forest | times | however | independence |
| Topic 7 | Topic 8 | Topic 9 | Topic 10 | Topic 11 | Topic 12 |
| using | oil | surveying | flood | geospatial | estate |
| marine | spill | laser | deformation | housing | real |
| area | change | sdi | disaster | government | government |
| malaysia | efficient | users | farm | without | owners |
| valuation | geoid | learning | consolidation | cost | market |
| surface | sustainable | traffic | realtime | digital | national |
| geospatial | datum | heritage | housing | squatter | economy |
| urban | levels | baseline | tax | act | delay |
| underground | assessment | terrestrial | correct | credit | due |
| monitoring | sea | university | buildings | performance | informal |

papers were published between January 1, 2008, and October 31, 2017, and included various cadastral research topics. I downloaded the abstracts of these articles and stored them in txt file format on a yearly basis. Through this process, I created a total of 10 documents. And I integrated ten documents into one corpus using R.

In this study, I analyzed the corpus using the following methods. First, I pre- processed the corpus using R. In other words, I transformed all words into lowercase letters and removed unnecessary symbols or punctuation.

Second, I created a document-term-matrix and changed the row names of this matrix to filenames. Third, I conducted LDA analysis using R topicmodels package to detect research topics and research trends in international cadastral research. I set the parameters for Gibbs sampling as follows. I set the burnin to 4000, the iteration to 1500, the thin to 500, and the nstart to 5 for an efficient topic identification.

Finally, I set the number of topics to 12, then use the topicmodels package to obtain per-corpus topic distributions, per-document topic proportions, and per-word topic assignments.

## 5. Results

### 5.1. Topics of Cadastral Researchers

Research Question 1 was about what topics cadastral researchers are most interested in. To answer this research question, I used LDA to obtain per-word topic assignments. Table 1 shows per-word topic assignments. As can be seen in this table, I set the number of topics to 12 based on the literature review results.

Topic 1 seems to be a topic closely related to professional bodies and education. In other words, the words assigned to this topic are mainly related to cadastre-related professional organizations and cadastral education. As can be seen in Table 1, the ten most probable words in this topic are professional, employers, membership, bodies, built, australia, employees, body, education, and students.

Topic 2 looks like a topic related to information systems. The words assigned to this topic are closely related to cadastral information systems and data model. As shown in Table 1, the terms that play a vital role in this topic are found to be data, system, paper, information, study, model, based, spatial, used, property.

Topic 3 looks like a topic very closely related to energy and marine area landuse. In other words, this topic has been identified as a topic closely related to alternative energy development and land use and management in the sea and coastal areas.

The most influential terms in this topic turn out to be energy, marine, area, landuse, and so on. Topic 4 relates not only to cadastral investment but also to a cadastral database. In this topic, investment, cadastral, and cadastre are closely associated with cadastral investment, and modules, datasets, and domains are closely related to cadastral databases.

Topic 5 deals with land development and land administration. This topic is one of the most influential topics in the whole corpus. In this topic, land, development, building, and construction are closely related to land development, and administration, right, and system are terms related to land administration.

Topic 6 is a topic related to Nigerian land market research. The terms contained in this topic are closely related to those that classify and study the Nigerian land market. The words that play a key role in this topic are identified as nigeria, market, study, protection, active, loads, classification.

Topic 7 relates not only to marine area valuation but also to urban underground monitoring. In other words, this topic is a subject that is closely related to marine cadastre and 3D cadastre, on which many scholars have been focusing recently. In this topic, marine, area, and valuation are closely related to marine cadastre, and urban, underground and monitoring are closely associated with 3D cadastre.

Topic 8 is deeply related to disasters and sustainable cadastre. Today, disasters such as earthquakes, floods and hurricanes are occurring all over the world. The role of cadastre in the process of recovering from these disasters is absolute. This topic is related to the role of cadastre in the process of restoring oil spill and its damage. In this topic, oil, spill, change are words

related to disaster, and efficient, geoid, sustainable, datum, level are terms related to sustainable cadastre.

Topic 9 is a topic related to surveying and cadastral education. It is not too much to emphasize the importance of education at times when the environment surrounding the cadastre field is rapidly changing as it is today. The words that play key roles in this topic turn out to be surveying, laser, sdi, users, learning, university, and so on.

Topic 10 is a topic related to natural disasters and land management. When natural disasters such as floods occur, farmlands are flooded and houses are destroyed. Accurate land management is needed to restore these flooded farmlands and rebuild houses. Topic 10 addresses the role of land management in the process of recovering damages from natural disasters. In this topic, flood, deformation, and disaster are terms related to natural disasters, and farm, consolidation, realtime, housing, correct, building are words related to land management.

Topic 11 is not only about the government's housing policy, but also about geospatial data. Housing, government, and squatter are words related to government housing policy, and geospatial, digital, and performance are terms related to geospatial data.

Finally, Topic 12 deals with real estate market and national economy. In this topic, estate, real, owners, and market are terms related to the real estate market, and government, national, and economy are identified as words closely related to the national economy.

## 5.2. Distribution of Topics

Research Question 2 is about how topics are distributed throughout the corpus. Figure 2 shows the proportion of each subject in the corpus. As can be seen in Figure 2, topic 2 was the biggest contributor to the overall corpus. This topic occupies more than one-third of the whole corpus, so it has been identified as the hottest topic of all topics. Based on these results, it can be seen that the researchers are focusing the most attention on researching the cadastral information system. Topic 5 is the second most important topic in the whole corpus. This topic is a topic that occupies 28% of the whole corps and relates to land development and land administration. Today, land development projects are being actively carried out in any country. There are many side effects in the land development process. It is the duty of the land administration experts to overcome these side effects in a wise manner and to suggest a new development direction. On the other hand, all other topics, including topic 1, occupy less than 5% of the entire corpus, and I can see that these are topics that do not play a big role in the whole corpus.

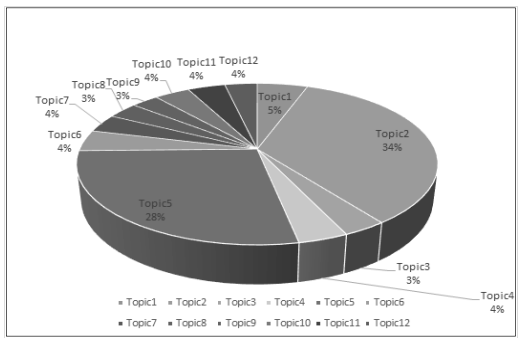


Figure 7. Per-corpus topic proportions

### 5.3. Influential Topics

Research Question 3 was about which topic was the most influential in each document. To answer these research questions, I used LDA to obtain per-document topic proportions. Table 2 shows the per-document topic proportions.

The most popular topic in 2008 was topic 1. In addition, topic 2 and topic 5 were found to be more influential than other topics. The topic that was the most influential in 2009 was topic 2. Topic 4 and topic 5 also turned out to be hot topics. This topic distribution pattern shows a big difference when compared with that of 2008. In 2010, researchers were most interested in cadastral information systems(topic 2).

Topic 5 and Topic 9 have also been identified as important topics in the overall document. This topic distribution pattern shows considerable similarity when compared with that of 2009.

The hottest topic of 2011 was also identified as topic 2. On the other hand, the most influential topic in the entire document in 2012 was topic 5. Also, topic 2 and topic 12 turned out to be important topics in the whole document. What is unique here, however, is the fact that topic 12 has entered the top spot for the first time.

The topic that was hottest in the entire document in 2013 turned out to be topic 6. As shown in table 2, the proportion of topic 6 in other years was not so high. Therefore, it is quite unusual that topic 6 occupies the largest portion of the whole document. Also, topic 2 and topic 5 were identified as hot topics in the entire document in 2013. The most influential topic in the whole document in 2014 was topic 2.

This topic occupies more than 40% of the entire document, making it the most influential topic in the entire document. The topic that was hottest in 2015 was topic 5. Also, topic 2 and topic 3 were found to be important topics in the overall document. What is noteworthy here is that topic 3

Table 2. Per-document topic proportions

| year | Topic1 | Topic2 | Topic3 | Topic4 | Topic5 | Topic6 | Topic7 | Topic8 | Topic9 | Topic10 | Topic11 | Topic12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2008 | 0.437769 | 0.343539 | 0.001652 | 0.010070 | 0.184001 | 0.003417 | 0.001652 | 0.005590 | 0.003010 | 0.002195 | 0.004368 | 0.002738 |
| 2009 | 0.017815 | 0.393619 | 0.007402 | 0.268351 | 0.262226 | 0.004952 | 0.008321 | 0.005564 | 0.008627 | 0.004339 | 0.008933 | 0.009851 |
| 2010 | 0.025236 | 0.413940 | 0.002102 | 0.026695 | 0.259709 | 0.006270 | 0.004082 | 0.008146 | 0.230113 | 0.006999 | 0.009396 | 0.007312 |
| 2011 | 0.003019 | 0.356837 | 0.007282 | 0.005577 | 0.322286 | 0.004937 | 0.009200 | 0.263463 | 0.004511 | 0.005364 | 0.013889 | 0.004084 |
| 2012 | 0.017263 | 0.284663 | 0.004459 | 0.008071 | 0.351964 | 0.018905 | 0.005608 | 0.004952 | 0.002325 | 0.018740 | 0.012995 | 0.270053 |
| 2013 | 0.005720 | 0.259129 | 0.00525 | 0.003840 | 0.325419 | 0.337643 | 0.013713 | 0.004780 | 0.009481 | 0.013713 | 0.011832 | 0.009481 |
| 2014 | 0.006534 | 0.429313 | 0.02412 | 0.022775 | 0.148047 | 0.012637 | 0.303524 | 0.016879 | 0.008293 | 0.012431 | 0.010672 | 0.004775 |
| 2015 | 0.006542 | 0.322453 | 0.259901 | 0.009691 | 0.340715 | 0.006752 | 0.010740 | 0.004023 | 0.010740 | 0.010320 | 0.002554 | 0.015568 |
| 2016 | 0.006218 | 0.266077 | 0.010268 | 0.008562 | 0.376714 | 0.007497 | 0.005578 | 0.007283 | 0.003446 | 0.296134 | 0.008562 | 0.003659 |
| 2017 | 0.013947 | 0.374401 | 0.008384 | 0.013251 | 0.205416 | 0.020901 | 0.003747 | 0.006993 | 0.008615 | 0.008384 | 0.317609 | 0.018351 |

occupies a large portion of the total document. In 2016, the largest topic in the whole document was topic 5. Topic 2 and topic 10 also turned out to be hot topics in the entire document. The topic that was hottest in 2017 was topic 2, and topic 5 and topic 11 were identified as topics that play an important role in the overall document. What is interesting here is that topic 11 has emerged as a hot topic. Watching the trends of topic 11 will be very interesting for cadastral researchers.

### 5.4. Changes of Topics over Time

Research Question 4 was about how cadastral research topics have changed over the years. To answer this research question, I used the per-document topic proportions of table 2 to create trendlines for the topics. Figure 3 shows cadastral research topic trends over time. As can be seen from the table, the very hot topics among the 12 topics were topic 2 and topic 5. Topic 2 is a topic related to cadastral information systems. This topic has attracted the attention of cadastral researchers over the past decade. As our society enters the information society, researchers in the field of cadastre have also become interested in the utilization of information systems. Also, Topic 5 has been a hot topic that has attracted significant interest from researchers over the last decade. This topic is a topic related to land development and land administration. As urbanization has rapidly advanced worldwide, many scholars have endeavored to create an efficient land management model that can minimize the adverse effects of land development. Thus, this topic attracted considerable attention from cadastral researchers every year. However, as can be seen in Figure 3, other topics did not draw the great interest of scholars except for a particular year. So they can be called cold topics. However, the fact that Topic 11 has emerged as a hot topic in 2017 is the point that we will pay attention to in the future.

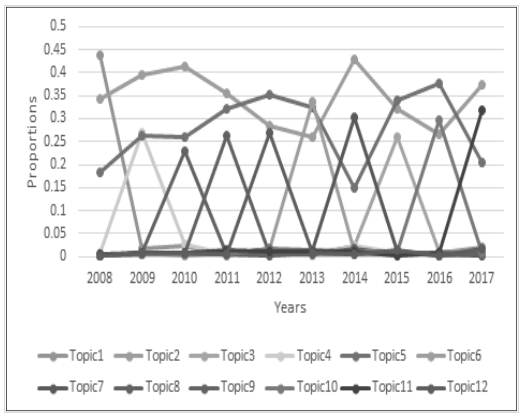


Figure 8. Topic trends over time

## 6. Discussion

### 6.1. Topics of Cadastral Researchers

The results of LDA analysis show that there are twelve topics in international cadastral research. However, as LDA analysis found duplicated topics, I classified them into six categories: cadastral information systems, land development and land administration, cadastral education, land market, disasters and sustainable cadastre, marine cadastre and 3D cadastre. First, topic 1 and topic 9 were found to belong to cadastral information systems category. This category is closely related to the study of technical aspects of cadastral research.

The importance of research on the technical aspects of cadastre has also been confirmed in previous studies (Paulson et al., 2011). It was also found that topic 5 and topic 10 belonged to the land development and land administration category. Topics in this category have played an important role in cadastral research(Enemark, 2009). The area of cadastral education has also attracted a great deal of attention from researchers in recent years. As the environment surrounding the cadastre changed drastically, the demand for education increased. It has been confirmed that topic 1 and topic 9 belong to this category. The results of this study are consistent with the results of previous studies (Manzano-Agugliaro et al., 2016). Many scholars are also focusing on land market. Topic 6 and topic 12 were found to belong to land market category. And researchers are concentrating on disasters and sustainable cadastre. Disasters continue to occur every year worldwide. Cadastre plays an important role in the process of restoring disaster damages. It was found that topic 8 and topic 10 belonged to this category. Recently, 3D cadastre and marine cadastre are becoming important topics. Especially, as the densification of urban land use accelerated, the importance of 3D cadastre increased. Topic 3 and topic 7 were found to belong to this category. As we have seen so far, most of the cadastre research topics derived from the LDA analysis closely coincide with previous research results. However, the legal aspect of the cadastre that has been confirmed as an important category in previous studies has not been identified as an important research topic in this study(Enemark, 2009).

### 6.2. Distribution of Topics

As our society enters the information society, the demand for information systems in all fields has increased sharply. The cadastral sector is no exception. The primary purpose of cadastre is to collect information related to a parcel of land and manage it efficiently. To achieve the objective of cadastre efficiently, governments around the world have tried to build cadastral information systems for a long time. Due to this realistic need and the aspirations of scholars, informatization in the cadastral field is accelerating. At the core of this cadastral computerization are cadastral information systems. The cadastral information systems are efficient systems that collect accurate data on a parcel of land and store and manage it. Due to this historical background, scholars' interest in cadastral information systems is higher than ever before. And this field is expected to lead the cadastral research in the future.

In the field of cadastral research, hot topics as much as cadastral information systems are land development and land administration. Today, as urbanization has progressed rapidly around the world, there are side effects caused by over-exploitation. It is essential to collect accurate land information and manage it efficiently to prevent such adverse effects in the land development process and to implement development projects more efficiently. In other words, active land management and administration are more important than ever. Today, many scholars are concentrating on the field of land development and land administration against this backdrop. However,

other topics accounted for less than 5% of the whole corpus, confirming that they did not attract the great interest of cadastre researchers. But, topic 11, i.e., government housing policy and geospatial data, has emerged as a hot topic in 2017, so it will be interesting to look at trends in this area in the future.

### 6.3. Influential Topics

The most influential topic in 2008 was topic 1. However, topic 1 did not have much influence except for 2008. This trend began to change in 2009. From this point on, the researchers became interested in the cadastral information systems (topic 2). In other words, scholars focused on how to build cadastral information systems and how to collect and manage geospatial data efficiently. This trend continued for three years until the end of 2011. But by 2012, this trend had begun to change. From this point on, researchers began to pay attention to the land development and land administration(topic 5). This change in trends was closely related to the large-scale urban development projects that were conducted around the world at the time. By the year 2013, topic 6 has become the hot topic. However, topic 6 did not have much influence except for 2013. By the year 2014, topic 2 was once again in the lead, but this lead did not last long. By the year 2015, topic 5 was the most influential topic in the whole document. This trend continued until the end of 2016. However, by the year 2017, topic 2 again took the lead in the whole document. As can be seen above, over the past decade, scholars have concentrated their efforts mainly on researching cadastral information systems, land development and land administration. The results of this analysis will be helpful for predicting future trends in international cadastral research.

### 6.4. Changes of Topics over Time

The topics that have played a key role in international cadastral research over the past decade are topic 2 and topic 5. These topics have led research in the cadastral field each year. In particular, topic 2(cadastral information systems) has been identified as the most influential topic in the last decade. This is closely related to the global trend of informatization. Many researchers and practitioners have been longing for a shift away from traditional analog cadastre to a new three-dimensional digital cadastre. Their dreams came true thanks to the IT revolution. Globally, researchers have made efforts to digitize cadastral maps and build cadastral databases to overcome analog cadastral disadvantages. In particular, researchers have made great efforts to develop rational models of cadastral information systems and cadastral databases. The efforts of these researchers and practitioners are well reflected in Figure 3. Also, topic 5(land development and land administration) was a hot topic that has led international cadastral research for the past decade. Recently, land development projects have been being actively implemented all over the world. Providing accurate land information in this process was an important task of cadastral practitioners. In the land development process,

disputes often arise between the neighbors over land boundaries. In other words, differences in perceptions of boundaries are often responsible for disputes between neighbors. In this context, many researchers have tried to suggest a reasonable land administration model that can minimize land disputes. This trend is expected to continue.

## 7. Conclusion

The primary purpose of this study was to discover the topics and trends of international cadastral research. To achieve these research objectives, I analyzed the collected data using LDA. As a result of the analysis, I have identified twelve significant topics in international cadastral research over the past decade. The most prominent topic among these twelve topics was topic 2, which accounted for more than one-third of the corpus, and topic 5 was also identified as a very hot topic. Other topics, on the other hand, occupy less than 5% of the entire corpus, confirming that they do not play important roles in the corpus as a whole. Cadastral researchers have been working intensively on specific areas over the past decade. In other words, researchers focused their attention on cadastral information systems, land development and land administration. The researchers' interest in these topics has continued for the past 10 years. Thus, topic 2 and topic 5, whose trend lines moved vigorously, turned out to be hot topics and other topics, whose trend lines showed a stagnation, were identified as cold topics.

This study can contribute to cadastral research in several aspects. First, this study was the first attempt to classify the topics of international cadastral research using LDA analysis. That is, in this study, I used LDA so that I could get a better understanding of the research topics. Second, this study was the first attempt to analyze the research trends in international cadastral research scientifically. Using the per-document topic proportions, I can see how the key topics change by year, i.e., by a document. Finally, using LDA will help researchers to grasp the topics and research trends of domestic and international cadastral research in the future. Despite its usefulness and applicability, this study has the following limitations. First, this study analyzed only the papers in the FIG Peer Review Journal to identify the topics and trends of international cadastral research. International cadastral journals include not only the FIG Peer Review Journal, but also journals published by ICA and IPRS. Second, in this study, I used only LDA to identify the topics and trends of cadastral research. However, LDA cannot detect the relationship between the terms in a topic. Finally, I only discovered the topics and research trends of international cadastral research.

To overcome these limitations, I will carry out the following research. First, I will analyze not only the papers published in the FIG Peer Review Journal but also the journals published in ICA and IPRS in future cadastral studies. Doing so will allow researchers to identify international cadastral research topics and trends more objectively. Second, I will use social network analysis as well as LDA to detect the relationships between the key

terms in a topic. Finally, I will compare and analyze the topics and trends of cadastral research using not only foreign materials but also domestic cadastral research papers.

## 참고문헌
## References


Bisnath, S., Saeidi, A., Wang, J. G., & Seepersad, G. (2013). Evaluation of Network RTK Performance and Elements of Certification—A Southern Ontario Case Study. Geomatica, 67(4), 243-251.

Blei, D. M. (2012). Probabilistic topic models. Communications of the ACM, 55(4), 77-84.

Blei, D. M., & Lafferty, J. D. (2006, June). Dynamic topic models. In Proceedings of the 23rd international conference on Machine learning (pp. 113-120). ACM.

Blei, D. M., & Lafferty, J. D. (2009). Topic models. Text mining: classification, clustering, and applications, 10(71), 34.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. Journal of machine Learning research, 3(Jan), 993-1022.

Braam, R. R., Moed, H. F., & Van Raan, A. F. (1991). Mapping of science by combined co-citation and word analysis I. Structural aspects. Journal of the American Society for inBlei, D. M., Ng, A. Y., & Jordan, M. I. (2003).

Boyack, K. W., & Klavans, R. (2010). Co - citation analysis, bibliographic coupling, and direct citation: Which citation approach represents the research front most accurately?. Journal of the American Society for Information Science and Technology, 61(12), 2389-2404.

Chang, J., Gerrish, S., Wang, C., Boyd-Graber, J. L., & Blei, D. M. (2009). Reading tea leaves: How humans interpret topic models. In Advances in neural information processing systems (pp. 288-296).

Chaney, A. J. B., & Blei, D. M. (2012, March). Visualizing Topic Models. In ICWSM.

Chen, C., & Carr, L. (1999, February). Trailblazing the literature of hypertext: author co-citation analysis (1989 - 1998). In Proceedings of the tenth ACM Conference on Hypertext and hypermedia: returning to our diverse roots: returning to our diverse roots (pp. 51-60). ACM.

Choi, H. S., Lee, W. S., & Sohn, S. Y. (2017). Analyzing research trends in personal information privacy using topic modeling. Computers & Security, 67, 244-253.

Das, S., Sun, X., & Dutta, A. (2016). Text Mining and Topic Modeling of Compendiums of Papers from Transportation Research Board Annual Meetings. Transportation Research Record: Journal of the Transportation Research Board, (2552), 48-56.

Deininger, K., Selod, H., &Burns, A. (2012). The Land Governance Assessment Framework: Identifying and monitoring good practice in the land sector. World Bank Publications.

Ding, Y., Chowdhury, G., & Foo, S. (1999). Mapping the intellectual structure of information retrieval studies: an author co-citation analysis, 1987-1997. Journal of information science, 25(1), 67-78.

Enemark, S. (2009). Sustainable Land Administration

Infrastructures to support Natural Disaster Prevention and Management. In The United Nations Regional Cartographic Conference for the Americas.

Estabrooks, C. A., Derksen, L., Winther, C., Lavis, J. N., Scott, S. D., Wallin, L., & Profetto-McGrath, J. (2008). The intellectual structure and substance of the knowledge utilization field: A longitudinal author co-citation analysis, 1945 to 2004. Implementation Science, 3(1), 49.

Franco, J. C. (2010). Contemporary Discourses and Contestations around Pro - Poor Land Policies and Land Governance. Journal of Agrarian Change, 10(1), 1-32.

Gatti, C. J., Brooks, J. D., & Nurre, S. G. (2015). A historical analysis of the field of or/ms using topic models. arXiv preprint arXiv:1510.05154.

Gregoire, D. A., Noel, M. X., Déry, R., & Béchard, J. P. (2006). Is there conceptual convergence in entrepreneurship research? A co - citation analysis of frontiers of entrepreneurship research, 1981 – 2004. Entrepreneurship theory and practice, 30(3), 333-373.

Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. Proceedings of the National academy of Sciences, 101(suppl 1), 5228-5235.

Hansen K.D., Gentry J., Long L., Gentleman R., Falcon S., Hahne F., & Sarkar D. (2017). Rgraphviz: Provides plotting capabilities for R graph objects. R package version 2.20.0.

Gučević, J., & Miljković, S. (2014). Testing the Operational Quality of the Permanent GNSS stations network within AGROS RTK Service. Geonauka, 2(1), 14-19.

He, Y., & Hui, S. C. (2002). Mining a web citation database for author co-citation analysis. Information processing &management, 38(4), 491-508.

Hoffman, M., Bach, F. R., & Blei, D. M. (2010). Online learning for latent dirichlet allocation. In advances in neural information processing systems (pp. 856-864).

Hsiao, C. H., & Yang, C. (2011). The intellectual development of the technology acceptance model: A co-citation analysis. International Journal of Information Management, 31(2), 128-136.

Karpik, A. P., & Musikhin, I. A. (2016). RESEARCH AND PRACTICAL TRENDS IN GEOSPATIAL SCIENCES. International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences, 41.

Kim, Y. K. & Lee, S. B. (2006). A Study on Improving Cadastre System and Cadastre-related Laws for Introducing 3D Cadastre. The Korea Spatial Planning Review, 2006.9, 147-162.

Lafferty, J. D., & Blei, D. M. (2006). Correlated topic models. In Advances in neural information processing systems (pp. 147-154).

Liu, Y., Niculescu-Mizil, A., & Gryc, W. (2009, June). Topic-link LDA: joint models of topic and author community. In proceedings of the 26th annual international conference on machine learning (pp. 665-672). ACM.

Manzano-Agugliaro, F., Castro-García, M., Pérez-Romero, A. M., García-Cruz, A., Novas, N., & Salmerón-Manzano, E. (2016). Alternative methods for teaching cadastre and remote sensing. Survey

Review, 48(351), 450-459.

Mcauliffe, J. D., & Blei, D. M. (2008). Supervised topic models. In Advances in neural information processing systems (pp. 121-128).

Neis, P., & Zielstra, D. (2014). Recent developments and future trends in volunteered geographic information research: The case of OpenStreetMap. Future Internet, 6(1), 76-106.

Nerur, S. P., Rasheed, A. A., & Natarajan, V. (2008). The intellectual structure of the strategic management field: An author co-citation analysis. Strategic Management Journal, 29(3), 319-336.

Paulsson, J., & Paasch, J. (2011). 3D property research-a survey of the occurrence of legal topics in publications. In 2nd International Workshop on 3D Cadastres, Delft, The Netherlands, 16-18 November, 2011 (pp. 1-14). International Federation of Surveyors.

Paulsson, J., & Paasch, J. M. (2013). 3D property research from a legal perspective. Computers, Environment and Urban Systems, 40, 7-13.

Paulsson, J., & Paasch, J. M. (2015). The land administration domain model-a literature survey. Land use policy, 49, 546-551.

Rizos, C. (2007). Alternatives to current GPS-RTK services and some implications for CORS infrastructure and operations. GPS solutions, 11(3), 151-158.

Ramage, D., Hall, D., Nallapati, R., & Manning, C. D. (2009, August). Labeled LDA: A supervised topic model for credit attribution in multi-labeled corpora. In Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1-Volume 1 (pp. 248-256). Association for Computational Linguistics.

Schildt, H. A., Zahra, S. A., & Sillanpää, A. (2006). Scholarly communities in entrepreneurship research: a co-citation analysis. Entrepreneurship Theory and Pra.

Schwarz, C. (2017). ldagibbs: A command for Topic Modeling in Stata using Latent Dirichlet Allocation. Forthcoming Stata Journal.ctice, 30(3), 399-415.

Silva, L. (2007). Institutionalization does not occur by decree: Institutional obstacles in implementing a land administration system in a developing country. Information Technology for Development, 13(1), 27-48.

Silva, M. A., & Stubkjær, E. (2002). A review of methodologies used in research on cadastral development. Computers, Environment and Urban Systems, 26(5), 403-423.

Steudler, D., Rajabifard, A., &Williamson, I. P. (2004). Evaluation of land administration systems. Land Use Policy, 21(4), 371-380.

Sun, L., & Yin, Y. (2017). Discovering themes and trends in transportation research using topic modeling. Transportation Research Part C: Emerging Technologies, 77, 49-66.

van Oosterom, P. (2013). Research and development in 3D cadastres. Computers, Environment and Urban Systems, 40, 1-6.

Wang, C., & Blei, D. M. (2011, August). Collaborative topic modeling for recommending scientific articles. In Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 448-456). ACM.

Watkins, D., & Reader, D. (2002). Identifying current trends in entrepreneurship research: A new approach. ARPENT: Annual Review of Progress in Entrepreneurship, 2, 311.

White, H. D., & McCain, K. W. (1998). Visualizing a discipline: An author co-citation analysis of information science, 1972-1995. Journal of the American society for information science, 49(4), 327-355.

Williamson, I., Enemark, S., Wallace, J., &Rajabifard, A. (2010). Land administration for sustainable development (p. 487). Redlands, CA: ESRI Press Academic.

Williamson, I., & Ting, L. (2001). Land administration and cadastral trends—a framework for re-engineering. Computers, Environment and Urban Systems, 25(4), 339-366.

Zhu, S. (2014). Pain expression recognition based on pLSA model. The Scientific World Journal, 2014.

2018년 4월 03일 원고접수(Received)
2018년 6월 01일 1차심사(1st Reviewed)
2018년 6월 15일 2차심사(2st Reviewed)
2018년 6월 27일 게재확정(Accepted)

**초 록**

본 연구의 주된 목적은 LDA를 이용하여 국제지적연구의 주제와 연구추세를 확인하는 것이었다. 이러한 연구목적을 달성하기 위해 나는 LDA와 국제지적연구에 관한 선행연구를 검토하였고 이를 기반으로 4 개의 연구 질문을 설정하였다. 이러한 연구 질문에 답을 구하기 위해 나는 FIG Peer Review Journal에 2008년 1월1일 부터 2017년 10월 31일 사이에 발표된 370편의 논문들을 LDA를 이용하여 분석하였다. 분석의 결과 나는 국제지적연구에 12개의 주요 주제가 존재하고 있음을 확인하였다. 그리고 이러한 주제 중에 가장 영향력 있는 주제는 topic 2 (지적정보시스템)로 확인되었으며 또한 topic 5 (토지개발과 토지행정)도 전체 문서에서 중요한 역할을 수행하고 있는 주제로 파악되었다. 이 두 주제는 지난 10년 동안 추세선이 매우 활발하게 움직인 가장 인기 있는 주제들로서 앞으로의 지적연구에서도 주도적인 역할을 수행할 것이 틀림없다.

주요어 : 주제, 추세, LDA, 국제지적연구, 지적정보시스템, 토지개발, 토지행정