

# 텍스트마이닝과 네트워크 분석을 활용한 미래예측 방법 연구

A Study of foresight method based on textmining  
and complexity network analysis

정 근 하

한국과학기술기획평가원



## 제 출 문

한국과학기술기획평가원장 귀하

본 보고서를 “텍스트마이닝과 네트워크 분석을 활용한 미래 예측방법 연구”의 최종보고서로 제출합니다.

2010년 12월

연구기관 : 한국과학기술기획평가원

연구책임자 : 정근하(선임연구위원)

참여연구원 : 정철우(위촉연구원)



## 요 약 문

본 연구에서는 미래예측 시 전문가들의 정성적인 의견과 평가를 보조할 수 있는 정량적이고 객관적인 자료를 도출하기 위하여 인터넷과 네트워크 기법 및 텍스트마이닝 기법을 활용한 방법을 제시하고자 하였다. 그리고 이러한 복합기법을 이용하여 건설분야를 대상으로 미래예측에 필요한 키워드를 도출하여 우선순위를 분석하였다. 특히, 건설분야의 분석대상은 ‘제3회 과학기술예측조사 수정·보완’(2008)에서 분석한 건설분야의 기술과 ‘과학기술 미래비전’(2010)을 대상으로 미래예측에 필요한 키워드들을 분석하였다. 또한, 2개의 모집단에서 분석되어진 키워드들은 정부에서 2007년에 발표된 ‘건설교통 R&D 중장기계획(2008~2012)’(2007)과 비교하여 건설분야의 기술개발 방향을 살펴보았다.

정성적인 방법에 의해 작성된 제3회 과학기술예측조사(대조군)의 기술들과 정량적인 방법(텍스트마이닝)에 의한 과학기술 미래비전(실험군)의 기술들을 비교분석한 결과, 정량적인 방법(텍스트마이닝)에 의해 도출된 기술들은 전문가의 정성적인 방법에 의한 도출된 기술들과 별로 차이가 없는 것으로 나타났다. 그러므로 본 연구에서 사용된 텍스트마이닝 기법은 미래예측 키워드를 도출하는 정량적인 방법으로 유용할 것으로 증명되었으며, 전문가들의 정성적인 자료를 보조하는 정량적인 자료분석 방법으로 유용할 것으로 기대된다. 또한, 텍스트마이닝 대상 자료의 수준과 종류에 따라 오히려 더 좋은 결과를 도출하는 하나의 방법으로 더 유용할 것으로 예상된다.

인터넷을 활용한 예측기법은 인터넷의 복잡하고 다양한 정보를 활용하여 현재의 트렌드를 반영할 수 있는 유용한 수단이며, 최근 페이스북, 트위터 등 소셜 네트워크(Social Network)를 활용한 인터넷은 스마트폰 등의 하드웨어적인 요소와 인간의 아날로그적인 감성을 활용한 수단으로 엄청나게

빠른 속도로 진화하고 있다. 그러므로 향후, 인터넷의 검색결과를 활용한 방법과 트렌드를 분석할 수 있는 텍스트마이닝 방법에 대한 다양한 연구가 필요하다. 또한, 복잡계의 다양한 방법들은 현재에도 진화하고 있기 때문에 어떻게 활용할 것인가에 대한 지속적인 연구가 필요할 것이다.

## SUMMARY

In this study, we present methods utilizing internet, network analysis, and text mining techniques in order to draw objectively quantified information that support experts' qualitative opinions and evaluations in forecasting. Furthermore, by applying this fabricated procedure, we have derived keywords to analyze priorities in architectural engineering. In particular, the target of this analysis is the architectural engineering technologies analyzed from "The 3rd Science and Technology Foresight" (2008) and the keywords drawn from "S&T Vision for the Future towards 2040" (2010) for forecasting. By comparing these two sets of analyzed keywords with the results from "Long-Mid Term Plan for R&D Activities of the Construction & Transportation (2008~2012)" (2007), we have investigated the research and development trends in the field of architectural engineering

Not much difference between expert's qualitative method and quantitative one such as text mining has been observed from the comparison study between the technologies derived from "The 3rd Science and Technology Foresight" qualitatively (control group) and the technologies from "S&T Vision for the Future towards 2040" quantitatively (experimental group). Therefore, as a useful quantitative tool for drawing keywords for forecasting, our text mining technique will support expert's quantitative analysis. In addition, depending on the raw data level and type, the text mining technique will bring a better result in forecasting keywords

derivation.

An internet forecasting method is useful when it utilizes the complex and diverse information and reflects the current trends. Recently, with the hardware components like smart-phones and the human sensitivities, internet has evolved with extreme rapidity as seen in social networks like Facebook and Twitter. Therefore, research activities accommodating internet search results and developing text mining methods for analyzing the current trend are in demand. Moreover, the question of using evolving methods in complex system only can be answered via continuous research activities.



# CONTENTS

Chapter 1 Introduction .....	1
1. Meaning & Purpose .....	1
2. Scope & Method .....	3
Chapter 2. Case Study .....	6
1. Outline .....	6
2. Qualitative Method .....	7
3. Quantitative Method .....	8
4. New Trend .....	10
Chapter 3. New Foresight Method .....	14
1. Complexity .....	14
2. Network Analysis .....	25
3. Network Analysis Method .....	31
4. Analysis using the Search Engine .....	37
5. Textmining .....	39
6. Information Extract of Textmining .....	42
7. Trend Analysis using the Textmining .....	50
8. New Complex Foresight Method .....	57
Chapter 4. Apply to Construction Field .....	61
1. The 3rd S&T Foresight .....	61
2. S&T Vision for the Future toward 2040 .....	72
3. Comparison Analysis of New Foresight Method .....	79
Chapter 5. Conclusion .....	100
<References> .....	104
<Appendix> 1. Textmining Software .....	108
2. Core Technology of Construction .....	116

## 목 차

제1장 서론 .....	1
제1절 연구의 필요성 및 목적 .....	1
제2절 연구의 범위 및 방법 .....	3
제2장 기존의 방법론에 대한 고찰 .....	6
제1절 개요 .....	6
제2절 정성적 방법 .....	7
제3절 정량적 방법 .....	8
제4절 새로운 동향 .....	10
제3장 새로운 미래 예측방법 개발 .....	14
제1절 복잡계의 개념 .....	14
제2절 네트워크 분석 .....	25
제3절 네트워크 분석방법 .....	31
제4절 검색엔진을 이용한 분석방법 .....	37
제5절 텍스트마이닝 .....	39
제6절 텍스트마이닝의 정보추출 방법 .....	42
제7절 텍스트마이닝을 이용한 트렌드 분석방법 .....	50
제8절 새로운 복합적 미래예측 방법 .....	57
제4장 건설분야에의 시범적용 .....	61
제1절 제3회 과학기술예측조사 .....	61
제2절 과학기술 미래비전 .....	72
제3절 새로운 예측방법과의 비교분석 .....	79
제5장 결론 및 시사점 .....	100
<참고문헌> .....	104
<별첨> 1. 텍스트마이닝 관련 소프트웨어 .....	108
2. 건설분야의 중점 과학기술 개요 .....	116

# 표 목 차

<표 1> 주요 예측방법의 장단점 비교 .....	31
<표 2> 네트워크 최적화 방법의 개념 및 장단점 비교 .....	3
<표 3> TF-IDF와 TF-Df 방식 비교 .....	1 5
<표 4> TF-IDF와 TF-Df 방식의 장단점 분석 .....	2 5
<표 5> 중점과학기술 세부기술 목록 및 개요 .....	3
<표 6> 기술 리스트 및 영문 키워드 .....	6
<표 7> 중요도 지수로(TI)로 도출된 기술 리스트 및 영문 키워드 .....	86
<표 8> 최종 도출된 기술 리스트 및 영문 키워드 .....	0
<표 9> 시나리오에 의한 주요 기술목록 .....	3
<표 10> 주요기술의 TF-Df .....	57
<표 11> 주요 기술의 우선순위에 따른 TF-Df .....	6 7
<표 12> 제3회 과학기술예측조사 텍스트마이닝 결과 .....	28
<표 13> 미래비전 텍스트마이닝 결과 .....	48
<표 14> 텍스트마이닝단계에서 도출된 키워드 .....	58
<표 15> 제3회 과학기술예측조사결과의 연결중심성 분석 .....	88
<표 16> 제3회 과학기술예측조사결과의 매개중심성 분석 .....	88
<표 17> 제3회 과학기술예측조사결과의 근접중심성 분석 .....	98
<표 18> 과학기술 미래비전결과의 연결중심성 분석 .....	98
<표 19> 과학기술 미래비전결과의 매개중심성 분석 .....	09
<표 20> 과학기술 미래비전결과의 근접중심성 분석 .....	9
<표 21> 예측조사와 미래비전의 비교분석 결과 .....	19
<표 22> 중장기계획(안)에서 우선순위가 높은 기술분야 .....	29
<표 23> 예측조사/미래비전/중장기계획(안)의 비교분석 결과 .....	49
<표 24> 실험군과 대조군의 비교분석 결과 .....	69
<표 25> 텍스트마이닝과 네트워크 분석의 복합기법 검증결과 .....	01

## 그림 목차

<그림 1> 텍스트마이닝을 활용한 프로세스 개요도 .....	4
<그림 2> 네트워크 분석을 위한 프로세스 .....	4
<그림 3> 연구방법과 프로세스 .....	5
<그림 4> 복잡계의 분류방법 .....	0
<그림 5> 스타형, Y형, 체인형, 써클형의 네트워크 .....	0
<그림 7> Random Network and Scale-Free Network .....	30
<그림 8> 패스파인더 예시 .....	5
<그림 9> 구글 검색엔진의 활용방법 .....	9
<그림 10> 텍스트마이닝 프로세스 .....	24
<그림 11> 검색엔진을 이용한 검색결과값 .....	55
<그림 12> 년도별 키워드의 절대적 빈도수 .....	75
<그림 13> 새로운 복합적 미래예측 방법의 과정 .....	06
<그림 14> 건설·교통분야 핵심기술리스트 도출과정 .....	63
<그림 15> 건설분야 네트워크 분석-I .....	17
<그림 16> 건설분야 네트워크 분석-II .....	27
<그림 17> 주요 기술의 년도별 변화 추이 .....	77
<그림 18> 주요 기술의 네트워크 분석 .....	87
<그림 19> 미래예측을 위한 키워드 분석 프로세스 .....	08

## 제1장 서론

### 제1절 연구의 필요성 및 목적

우리나라의 국가연구개발사업은 양적·질적으로 규모가 확대되고 있으며, 투자된 성과 대비 결과물의 중요성과 그 과정에서 이루어지는 프로세스들의 다양성이 중요한 요소가 되고 있다. 뿐만 아니라, 국가경쟁력을 강화하기 위하여 국가 과학기술의 미래를 전망하거나 기술을 예측하는 것이 중요한 이슈로 떠오르고 있으며, 이러한 미래전망을 통해 국가의 다양한 전략과 정책을 수립하고 있다.

국가연구개발사업이나 국가적으로 중요한 핵심기술들은 기술의 중요도, 시급성, 파급효과 등의 다양한 기준에 따라 전문가 평가에 의해 우선순위가 도출되고, 도출된 사업이나 과제들은 국가의 정책 및 전략에 따라 우선순위가 결정되어 다양한 형태로 시행되고 있다. 또한, 글로벌화가 진전되고 연구개발 분야와 규모가 커지면서 국가 경쟁력 차원에서 과학기술은 매우 중요한 역할을 담당하게 되었다. 특히, 미래를 예측하여 중요 과학기술들을 선별 및 발굴하는 작업은 중요한 이슈가 되고 있으며, 이와 관련된 정부부처의 역할과 연구들이 중요한 테마가 되는 실정이다. 또한, 정부 및 단체에서는 한정된 자원과 인력을 어떻게 투자 및 관리할 것인가에 대해서 미래예측과 관련하여 많은 관심이 집중되고 있다.

이러한 미래를 예측하는 방법들은 대부분 전문가의 정성적인 의견과 평가에 의해 이루어지고 있으며, 객관적인 방법론들에 대한 연구와 투자는 아직까지 초기단계에 머무르고 있는 실정이다. 그 중 전문가에 의한 정성적인 미래예측방법은 매우 중요한 요소로 인정되어 왔으나, 전문가들의 경향, 정치적 요소, 인맥 관계 등의 다양한 요소에 의해 일부 편향적인 의견이나 주

장이 반영되고 있는 실정이며, 또한, 전문가들의 평가도 객관적인 자료가 부족하여 미래를 예측할 경우 오류와 많은 어려움을 겪고 있는 실정이다.

따라서, 과학기술분야의 미래예측 시 전문가의 객관적인 의견과 평가가 이루어지도록 좀 더 구체적이고 객관적인 데이터와 자료를 제공하는 방법이 중요한 이슈로 떠오르고 있으며, 특히 기존의 논문, 특허 등 다양한 소스를 어떻게 객관화할 것인지에 대한 방법과 연구가 활발히 진행되고 있다. 본 연구에서는 미래예측 시 전문가들의 정성적인 의견과 평가를 보조할 수 있는 정량적이고 객관적인 자료를 도출하기 위하여 인터넷과 네트워크기법 및 텍스트마이닝기법을 활용한 방법을 제시하고자 한다. 텍스트마이닝과 복잡계 네트워크 분석을 통한 정량적인 미래예측 방법은 특정한 상황이나 조건을 반영하지 못하는 불연속적인 변화추세와 현재의 트렌드를 반영하지 못하는 정량적인 방법을 보완하였다. 특히, 실시간으로 변화되어진 트렌드를 반영하여 특정한 상황이나 조건하에서 주요 기술이나 키워드들의 우선순위를 도출할 수 있도록 개발되었다. 또한, 기존의 정성적인 예측방법에서 보여지는 전문가의 편향적인 미래예측이나 트렌드를 반영하기 어려운 기존의 방법을 보완하는 하나의 방법으로 제시하였다.

제시된 모델은 상호 연관성이 없는 과학기술들을 특정한 조건이나 상황에 연결되어 있다고 가정하여, 인터넷을 활용한 가상의 네트워크를 구성하는 방법이 핵심이다. 즉, 인터넷 검색엔진에 의해 검색되어지는 두개의 특정 키워드에 대한 검색 결과값을 해당 두개의 키워드가 인터넷에 노출되어지는 빈도수로 가정하고 그 검색 결과값을 상호연관관계 가중치로 활용하는 방법이다. 이러한 방법은 연결고리가 없는 노드들의 가상적인 네트워크를 제공할 뿐만 아니라, 인터넷의 검색소스(논문검색사이트, 특허검색사이트, 사전사이트, 블로그, 일반웹페이지, 언어별, 국가별, 지역별, 특정분야 등)에 따라 같은 노드라도 원하는 조건에 따라 다양한 상호연관관계를 구성할 수 있

다. 또한, 특정한 상황, 조건, 기간 등을 제시하고 제시된 새로운 가상 네트워크를 구성할 수 있다는 장점이 있다. 이는 기존의 네트워크 구성시 관련된 연결고리만을 대상으로 분석되어지는 한계를 극복한 것이다.

## 제2절 연구의 범위 및 방법

미래예측 방법을 위한 조사 및 분석을 위해 IT분야에 활용되는 텍스트마이닝 분석방법과 복잡계 네트워크 분석방법, 그리고 인터넷 및 대용량 데이터 처리를 위한 오픈API와 AJAX 등의 프로그래밍 방법 및 적용될 건설분야의 이슈를 검토한다.

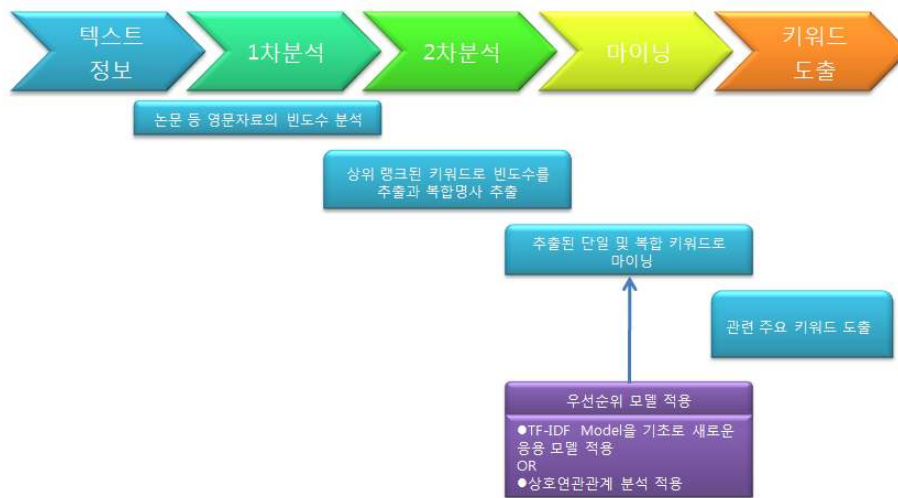
또한, 새로운 미래예측 방법을 연구하기 위해 텍스트마이닝을 통한 주요 키워드를 도출하고 분석한다. 텍스트마이닝은 워드프로세서, e-mail, 프리젠테이션, 스프레드시트, PDF와 같은 복합문서와 인터넷 페이지 등의 비정형화된 데이터(약 80%)를 대상으로 일정한 형식과 조건을 만족하는 자료로 가공하여 추출하는 일련의 방법이다.

본 연구에서는 건설분야 주요 저널지의 상위에 랭크된 논문을 대상으로 텍스트마이닝 방법을 활용해 주요 키워드를 도출하고 분석한다. 텍스트마이닝을 위한 주요 프로세스는 2단계에 걸쳐 분석할 예정이며, 1단계분석은 1~2차 분석을 통해 주요 논문의 선택적 방법과 그 논문들의 주요 키워드 분석, 2단계에서는 마이닝을 통한 주요 키워드를 도출한다.(<그림 1> 참조) 특히, 텍스트마이닝에서 주요 키워드의 도출을 위하여 TF-IDF<sup>1)</sup> 기본개념을 활용한 방법을 사용하거나 상호 연관관계의 분석 등 여러 가지 방법을 검토하여 최적의 방법을 선정한다.

---

1) TF-IDF(Term Frequency - Inverse Document Frequency)는 정보 검색과 텍스트 마이닝에서 이용하는 가중치로, 여러 문서로 이루어진 문서군이 있을 때 어떤 단어가 특정 문서 내에서 얼마나 중요한 것 인지를 나타내는 방법으로, 본 연구에서는 이 방식을 기본개념을 활용하여 응용하고자함.

#### 4 텍스트마이닝과 네트워크 분석을 활용한 미래예측 방법 연구



<그림 1> 텍스트마이닝을 활용한 프로세스 개요도

그리고 네트워크 분석을 통한 주요 키워드의 우선순위를 분석하기 위해 인터넷의 검색엔진을 활용하여 검색된 결과 값을 중심으로 키워드 간 상호연관관계의 중요도를 설정하고, Pathfinder를 통한 네트워크의 간소화와 네트워크 중심성 분석을 통한 중요도를 분석한다.



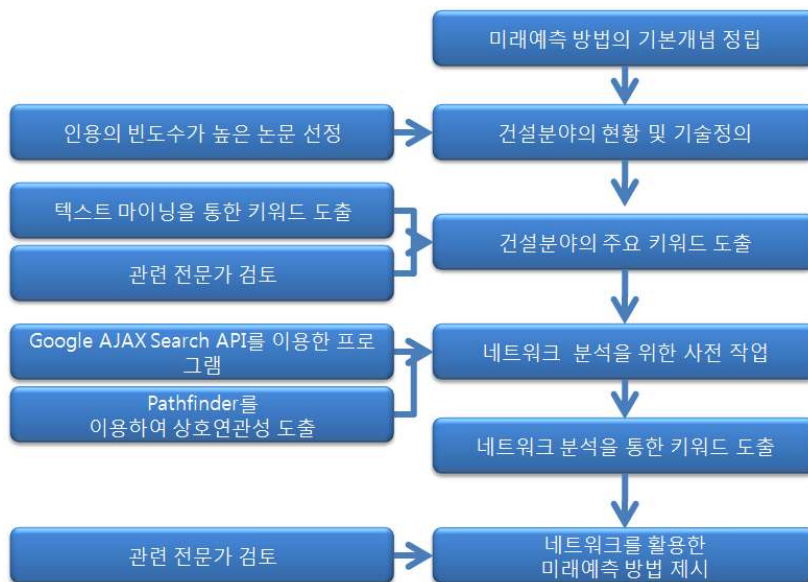
<그림 2> 네트워크 분석을 위한 프로세스



추진전략과 연구방법으로는 미래예측에 적합한 새로운 방법을 개발하기 위해 인터넷, 논문 등의 정보를 활용하여 객관적인 자료를 제시할 수 있는 미래예측 방법과 사회과학 방법론, 경제물리학 등의 통섭을 통한 다양한 방법을 검토한다. 이것은

미래예측 시 대부분 전문가에 의한 정성적인 방법에 객관적인 자료(정량적인 방법)를 제공할 수 있는 토대를 마련하고 기존의 전문가 활용이 어려운 경우나 시간과 노력이 많이 소요되는 전문가 평가를 일정부분 대체할 수 있다.

새로운 미래예측 방법을 위한 프로그램 개발의 토대를 마련하기 위해 공개 API, AJAX, 검색엔진 등의 IT를 활용하여 미래예측 방법의 효용성을 증대하고 대용량 및 다양한 데이터의 처리를 위한 프로그램 개발을 검토한다. 그리고 건설분야에 시범 적용하여 기존의 자료와 비교 검토하여 새로운 예측방법을 검증한다.



<그림 3> 연구방법과 프로세스

## 제2장 기존의 방법론에 대한 고찰

### 제1절 개요

과학기술의 발전은 우리의 사회, 문화, 경제, 정치 등 많은 분야에 영향을 미치고 있으며, 국가의 경제와 경쟁력에 미치는 영향은 그 어느 때보다도 증가하고 있다. 특히, 지구온난화, 에너지문제, 자원고갈 등과 같은 국가안보와 경제에 직접적으로 미치는 요소들은 해결하기 위하여 주요 선진국들은 막대한 예산을 투입하여 과학기술을 개발하고 있다.

따라서, 국가연구개발사업은 양적·질적으로 규모가 확대되고 있으며, 투자된 성과 대비 결과물의 중요성과 그 과정에서 이루어지는 프로세스들의 다양성이 중요한 요소가 되고 있다(최근 10년(1998~2007)동안 한국의 정부연구개발투자는 약 2.9배로 증가하였으며, 2008년도에는 미국, 일본, 프랑스, 독일, 영국, 이탈리아, 중국에 있어 세계 8번째로 정부연구개발투자의 지출규모가 10조원을 넘어서고 있음<sup>2)</sup>). 또한, 과학기술의 발전이 이러한 국가연구개발사업의 대상이 되는 사업이나 기술들은 각 분야의 전문가에 의해 평가 및 우선순위가 도출되고, 도출된 사업이나 과제들은 국가의 정책 및 전략에 따라 우선순위가 결정되어 다양한 형태로 시행되고 있다. 뿐만 아니라, 국가경쟁력을 강화하기 위하여 국가 과학기술의 미래를 전망하거나 기술을 예측하는 것이 중요한 이슈로 떠오르고 있다<sup>3)</sup>. 이러한 미래 변화의 트렌드를 파악하고 미래의 핵심기술을 선별하기 위하여, 주요 선진국들은 주기적으로 국가의 미래 트렌드를 분석하고 그 결과를 발표하고 있다<sup>4)5)6)7)</sup>.

---

2) 안승구, Analysis on the national R&D investment activities of Major countries in 2009 (KISTEP, 2009)

3) Michael Marien, Future Studies in the 21st Century: A Reality-Based. View, Futures, Vol.34 (2002) 261 - 281

이러한 미래 트렌드나 핵심기술을 선정하기 위하여 좀 더 객관적인 데이터와 자료를 제공하는 방법이 중요한 이슈로 떠오르고 있으며, 특히 기존의 논문, 특허 등 다양한 소스를 어떻게 객관화할 것인지에 대한 방법과 연구가 활발히 진행되고 있다<sup>8)9)</sup>.

## 제2절 정성적 방법

정성적인 방법중 대표적인 방법인 시나리오(1967, Kahn Herman)기법은 분류기준에 따라 다양한 미래의 모습을 보여주지만 의사결정을 위한 세부적인 내용은 잘 표현하지 못하므로, 시나리오 기법은 다른 기법들과 혼용하여 사용되어지고 있다. 이 외에도 정성적인 방법으로 내외부환경 분석 방법인 SWOT(1960년대, Albert Humphrey), 토론중에 얻은 지식이나 정보를 통하여 미래를 예측하는 전문가 패널(2004, Keenan), 일종의 브레인스토밍 과정의 구조화를 통한 기법인 미래마퀴(1971, Jerome Glenn), 나무의 가지처럼 미래의 유형을 구조화하는 기법인 의사결정나무(1964, Sonquist & Morgan) 등이 있다. 또한, 일반적으로 정략적인 통계기법을 사용하는 델파이(1949, RAND)는 전문가의 합의에 의해 불분명한 미래를 예측하는 방식으로 현재까지 가장 많이 활용되고 있으며, 이를 보완한 교차영향분석(1996, Gordon & Helmer), 리얼타임델파이(2006, Gordon & Pease) 등이 있다. 또한, 과거의 데이터를 이용하여 수세를 분석하는 추세

4) Governance in the 21st Century: FUTURE STUDIES(OECD, Paris, 2001)

5) Mikko Syrjänen, Yuko Ito, Eija Ahola (editors), Foresight for Our Future Society: Cooperative project between NISTEP(Japan) and Tekes(Finland)(Tekes & NISTEP, 2009)

6) Cornelia Daheim, Regional Foresight in Europe - 2 Examples: Duesseldorf and Linz, WFS Conference Minneapolis(2007)

7) Pirjo Ståhle (ed.), Five Steps for Finland's Future(Tekes, Helsinki, 2007)

8) Jeremy Ginsberg, Matthew H. Mohebbi, Rajan S. Patel, Lynnette Brammer, and Mark S. Smolinski & Larry Brilliant, Detecting influenza epidemics using search engine query data, Nature 457(2009) 1012-1014

9) Sang Hoon Lee, Pan-Jun Kim, Yong-Yeol Ahn, and Hawoong Jeong, Googling hidden interactions: Web search engine based weighted network construction, PACS(2008)

외삽법<sup>10)</sup>, 쌍대비교를 통한 우선순위를 도출하는 AHP(1970, Saaty) 등이 있다.

그러나, 위와 같은 정성적인 방법<sup>11)</sup>들은 사전정보가 부족한 미래를 예측할 수 있는 장점이 있는 반면 미래를 예측할 수 있는 사전정보의 부족과 전문가들의 경향, 정치적 요소, 인맥 관계 등의 다양한 요소에 의해 편향적인 예측결과를 초래하기도 한다(2007, 최항섭). 또한, 수학적, 통계적 기법을 활용하여 과거 데이터를 이용한 선형적 분석기법인 추세외삽법은 새로운 변수의 추가 및 다양한 변수에 의해 예측력이 떨어지고 특정한 상황 및 조건(불연속적인 변화)에 대하여 예측이 빗나가기 쉽다(2008, 이세준 외). AHP기법은 의사결정에 필요한 변수들을 1:1로 매칭시켜 상대적인 중요도를 평가하여 변수들의 우선순위를 판별할 수 있으나, 현재의 트렌드를 반영하는 것은 어렵다.

정성적인 방법은 일반적으로 브레인스토밍(1930, Alex Faickney Osborn), 전문가 패널(2004, Keenan), 미래바퀴(1971. Jerome Glenn), 환경스캐닝, 시나리오(1967, Kahn Herman)기법, 의사결정나무(1964, Sonquist & Morgan) 등이 있다. 이러한 방법들은 전문가의 지식과 의견에 따라 미래를 전망하는 방법이다.

### 제3절 정량적 방법

정량적인 방법에는 일반적으로 과거의 자료를 바탕으로 미래를 예측하는 기법인 추세외삽법[9]과 두 변수간의 쌍대비교를 통해 우선순위를 도출하

10) R. Levin, D. Rubin, and J. Stinson, Forecasting. in Quantitative approaches to management, Chapter 3(1986, NY, McGraw-Hill)

11) James A. Dator, Advancing Futures: Futures Studies in Higher Education(Praeger, Westport, 2002)

는 AHP(1970, Saaty) 등이 있다. 또한, 통계를 활용하는 정량적인 방법에는 델파이(1949, RAND)와 교차영향분석(1996, Gordon & Helmer)<sup>12)</sup>이 있으며, 이러한 기법들은 전문가의 합의에 의해 불분명한 미래를 예측하는 방식이다. 현재, 각 국가와 단체에서 가장 많이 사용하는 일반적인 과학기술 예측조사 방법은 델파이 조사와 시나리오 방법이다<sup>13)</sup>. 델파이조사는 1948년 미국의 국방연구소인 RAND연구소에서 개발되어 군사분야, 국가연구개발분야, 교육분야 등 다양한 분야에서 활용되고 있다. 델파이기법은 미래를 예측하는 경우 예측할 수 있는 자료가 부족하거나 거의 없을 때 전문가의 직관에 의해 합의점을 도출하는 방식을 말한다. 즉, 익명의 전문가들의 설문조사를 통해 반복적인 피드백을 통해 미래에 대한 자료를 수집하는 방법이다. 현재, 리얼타임델파이(2006, Gordon & Pease)가 등장하면서 다양한 형태의 델파이조사 방법<sup>14)</sup>이 개발되고 있다. 조사방법은 여러 전문가들의 서베이를 통해 어떤 기술의 실현시기를 주관적으로 예측하고, 예측한 값의 중앙값을 이용하여 예측하는 통계학적 방법이다. 대부분의 국가 미래 예측시 많이 사용하는 방법이다<sup>15)</sup>. 그리고, 시나리오기법<sup>16)</sup>은 전문가들에 의해 주어진 변수에 따라 다양한 미래의 모습을 예측하는 방식으로 서술식으로 작성되며 장기적인 미래의 모습을 담아낸다. 그러나, 이러한 방법은 거시적인 모습을 잘 설명하는 반면에 의사결정을 위한 세부적인 내용들은 기술되지 않는다. 시나리오에 의한 작성방법<sup>17)</sup>은 주어진 환경조건

12) 박병원, 과학기술예측조사를 위한 방법론 및 프레임워크 개선연구(KISTEP, 2007)

13) Hariolf Grupp and Harold A. Linstone, National Technology Foresight Activities Around the Globe: Resurrection and New Paradigms, Technological Forecasting and Social Change, Volume 60, Issue 1, 2 January(1999) 85-94.

14) Martin Hilbert, Ian Miles, and Julia Othmer, Foresight tools for participative policy-making in inter-governmental processes in developing countries: Lessons learned from the eLAC Policy Priorities Delphi, Technological Forecasting and Social Change, Volume 76, Issue 7(2009) 880-896

15) 손석호, Evaluation of Korean Technology Foresight Program(KISTEP, 2008)

16) Schoemaker, J.H. Pau, Scenario Planning: A Tool for Strategic Thinking, Sloan Management Review. Winter(1995) 25-40

17) Mats Lindgren, Hans Bandhold, Scenario planning: The link between future and strategy (Macmillan, 2002)

내에서 중심이 되는 이슈나 결정요인들을 파악하고, 이러한 요소들에 의해 다양한 미래의 시나리오를 작성한다.

#### 제4절 새로운 동향

이러한 미래예측을 위하여 인터넷을 활용한 객관적인 데이터를 활용하는 방법들은 다양하게 연구되어지고 있다. 특히, 검색엔진을 활용한 인터넷 활용방법은 시간 및 공간적인 제한 조건없이 분석이 가능하다는 장점이 있으며, 어떤 경우에는 전문가를 활용하지 않아 인력에 소비하는 노력과 시간을 절약할 수 있다. 또한, 인터넷의 검색소스(논문검색사이트, 특허검색사이트, 사전사이트, 블로그, 일반 웹페이지, 언어별, 국가별, 지역별, 특정분야 등)에 따라 같은 분야의 분석도 다양한 비교분석이 가능하다. 이러한 방법은 단독으로 사용하기보다는 다양한 기법과 결합되어 사용되어지는 것이 효용성측면에서 유용하다. 현재, 인터넷을 활용하는 방법은 다양하나, 본 연구와 관련된 유사한 방법은 다음과 같다.

최근 2008년 Nature에 발표된 논문[8]에 의하면 미국 시민들의 검색단어 빈도수를 이용하여 지역별 독감발생 추이를 분석한 결과, 미국 CDC(The U.S. Centers for Disease Control and Prevention)에서 발표한 실제 독감 발생 추이와 유사하고 이런 분석방법을 통해 CDC 발표시기보다 약 2주정도 빠른 결과를 도출할 수 있다. 또한, 2007년 구글 검색엔진을 활용하여 네트워크를 상호관계의 대한 가중치를 분석하는 방법도 연구되기 시작하였다.

또한, 요소들간 상호연관 관계를 의해 어떤 요소들이 중요한가에 대한 연구는 네트워크 연구에서 본격적으로 다루고 있다. 이러한 네트워크 이론은 1960년 수학분야의 그래프이론(graph theory)<sup>18)</sup>에서 본격적인 연구가

18) P. Erdős and A. Réney, Publ. Math. Inst. Hung. Acad. Sci. 5, 17(1960); Bull. Inst. Int. Stat. 38, 343(1961)

시작되었으며, 사회과학에서는 1930년대부터 사회 네트워크((Social network)라는 이론으로 수많은 연구가 이루어지고 있다. 네트워크 이론은 기본적으로 노드(점)와 링크(선)로 연결되어 있으며, 사회과학분야에서는 노드를 행위자로 링크를 상호연관관계로 취급하였다<sup>19)20)21)22)</sup>.

사회과학분야에서는 1940년 미국 MIT의 Alex Bavelas와 Harold Leavitt의 실험에 의해 네트워크는 중심(center)이라는 개념이 중요한 요소가 되었다. 특히, 중심에 위치하는 행위자는 집단내 파워가 강해지기도 하고(Brass, 1984), 개인적인 혁신을 용이하게 하며(Ibarra, 1993), 의사결정에 영향력을 행사할 수 있으며(Friedkin, 1993), 개인의 성과가 높아진다고(Baldwin, Bedell & Johnson, 1997) 한다. 또한, 한 기업에서의 자신의 네트워크의 위치에 따른 승진기회의 차이(Burt, 1992)가 생기며, 구조적으로 중용한 위치를 차지하고 있는 조직은 전략적으로 중요한 위치를 통한 정책적 함의를 이끌어 낼 수 있다(Scott, 2000) 그래프이론(graph theory)과 무작위 네트워크의 연결수의 분포함수<sup>23)</sup>에 대한 연구 이후, 1998년 Watts와 Strogatz는 ‘좁은세상 네트워크’라는 모형을 제시하고 전력망, 선충의 신경망, 영화배우 연결망을 통해 결집계수로 측정하였다<sup>24)</sup>. 이 모형에서 노드간 연결된 링크는 지름길을 통해 전체 네트워크가 긴밀해 진다는 설명하였다. 이 연구를 통해 네트워크의 연구가 폭발적으로 증가하기 시작하였으며, 척도없는 네트워크를 발표하게 되는 계기가 되었다. 1999년 R. Albert, H. Jeong, A.L. Barabasi에 의해 하이퍼링크로 연결된 인터넷의 물리적 연결구조에 대한 논문을 발표하였다<sup>25)</sup>. 이 연구결과로 연결수 분포

19) 윤영수, 채승병, 복잡계 개론(삼성경제연구소, 2005)

20) 손동원, 사회 네트워크 분석(경문사, 2002)

21) 김용학, 사회연결망 이론(박영사, 2004)

22) 강병남, 복잡계 네트워크 과학(집문당, 2009)

23) B. Bollobas, Discrete Math. 33(1981), 1

24) D.J Watts and S.H. Strogatz, Collective dynamics of ‘small-world’ networks, Nature 393(1998)

25) R. Albert, H. Jeong, and A.L. Barabasi, Diameter of the World-Wide Web, Nature 401(1999)

가 무작위 네트워크가 아니라, 멱함수법칙(power law)를 따르는 것으로 나타났다으며 많은 분야에서 이러한 현상들이 밝혀지게 되었다. 이를 척도없는 네트워크(scale-free network)라고 명명되었다. 2000년대 이후 네트워크는 사회과학에서 뿐만 아니라, 신진대사 반응 네트워크<sup>26)27)28)</sup>, 단백질 상호작용 네트워크<sup>29)</sup>, 논문인용분야<sup>30)31)</sup>, 주식시장 분석<sup>32)</sup>, 인터넷 구조분석<sup>33)</sup>, 전염병 연구<sup>34)</sup> 등 여러 분야에서 다양하게 응용되고 있다.

---

26) S. A. kauffman, The origin of order : Self-organization and selection in evolution(Oxford University Press, New York, Oxford)

27) H. Jeong, B. Tombor, R. Albert, Z.N. Oltvai, and A.L. Barabasi, Nature 407(2000)

28) A. Wagner and D.A. Fell, Proc. R. Soc. London B 268, 1803(2001)

29) M.E. J. Newman, SIAM Review 45(2003), 167

30) S. Redner, Eur. Phys. J. B 4(1998)

31) C. Tsallis and M.P. de Albuquerque, Eur. Phys. J. B13(2000)

32) Rosario N. Mantegna and H. Eugene Stanley, An Introduction to Econophysics: Correlations and Complexity in Finance(Cambridge University Press, Cambridge, 1999)

33) M. Faloutsos, P. Faloutsos, and C. Faloutsos, Comput. Commun. Rev. 29, 251(1999)

34) Liljeros et al, Nature(2001)



&lt;표 1&gt; 주요 예측방법의 장단점 비교

방법론	특징	장점	단점
델파이 기법	<ul style="list-style-type: none"> <li>- 주어진 예측과제에 대한 전문가들의 의견을 수렴하여 미래를 예측하는 방법</li> <li>- 통계 분석방법으로 대부분 국가에서 사용하고 있음.</li> </ul>	<ul style="list-style-type: none"> <li>- 자료 부족 및 거의 없을 때 미래예측 방법</li> <li>- 중앙값을 이용하여 예측시기 등이 수치로 도출</li> <li>- 익명성과 독립성으로 참여자의 자유로운 의견 개진</li> </ul>	<ul style="list-style-type: none"> <li>- 전문가 풀의 선정과 유지에 많은 시간과 노력이 필요</li> <li>- 최소 2회이상의 설문 조사를 인해 최소 6개월이상의 기간이 소모되며, 대단히 많은 비용이 소모됨.</li> <li>- 예측항목간 존재하는 상호 관계를 무시</li> </ul>
교차영향 분석	<ul style="list-style-type: none"> <li>- 델파이기법처럼 수치적인 결과를 도출하는 방법으로 예측 항목간 존재하는 상호 관계를 무시하는 델파이기법의 보완 기법</li> <li>- 시나리오를 작성하고 시나리오간 교차분석이 가장 많이 활용</li> </ul>	<ul style="list-style-type: none"> <li>- 예측항목간 존재하는 상호 관계를 구조화함</li> <li>- 상대적으로 계산이 용이하고 시각적으로 쉽게 파악 할 수 있음.</li> </ul>	<ul style="list-style-type: none"> <li>- 사전에 델파이조사를 실시 하여야함.</li> </ul>
시나리오 기법	<ul style="list-style-type: none"> <li>- 미래에 발생할 수 있는 개연성이 높고 영향력이 높은 시나리오를 제시하는 방법</li> <li>- 정성적인 분석방법으로 중장기 미래예측에 사용됨.</li> </ul>	<ul style="list-style-type: none"> <li>- 대안적인 미래 제시</li> <li>- 다른 기법하고 결합시 다양한 용도로 활용이 가능함.</li> </ul>	<ul style="list-style-type: none"> <li>- 세부적인 기술에 대한 접근이 용이하지 않음.</li> <li>- 의사결정과 실행을 위한 자료로써 미흡함.</li> </ul>
복잡계 가상 네트워크 모델	<ul style="list-style-type: none"> <li>- 인터넷과 복잡계 네트워크 분석 활용하여 핵심기술을 도출</li> <li>- 정량적인 분석방법</li> </ul>	<ul style="list-style-type: none"> <li>- 전문가로 활용하지 않으므로 많은 시간, 노력, 비용을 들지 않음.</li> <li>- 일부 전문가의 평가를 대처 할 것으로 예상됨.</li> </ul>	<ul style="list-style-type: none"> <li>- 델파이, 시나리오 등의 방법과 병행하면 효율적임.</li> </ul>

### 제3장 새로운 미래예측 방법 개발

#### 제1절 복잡계(Complexity)의 개념

현재 우리가 살아가는 세상에서 발생하는 복잡하고 다양한 자연 및 사회학적 현상을 이해하기 위하여, 창발되거나 서로 상호연관 작용하는 현상 등의 여러 가지 현상들을 접근하는 새로운 이해가 필요하게 되었다. 이러한 복잡한 세계를 이해하기 위하여, 20세기초부터 도입된 개념은 사회 및 경제현상 등 다양한 분야에서 연구가 진행되고 있었으며, 본격적인 연구는 1984년 미국 뉴멕시코주에 있는 산타페 연구소<sup>35)</sup>가 설립되면서 시작되었다. 이러한 복잡계는 수많은 복잡한 자연현상들이 특정한 상호작용에 의해 연결되어 있으며, 이러한 상호작용은 비선형적이며 특정한 관계가 없는 것처럼 보인다. 최근, 이러한 현상들을 일부 설명할 수 있는 자기유사성 즉 프랙탈, 자기조직화, 임계현상 등의 이론들이 제시되었으며, 이러한 이론들은 거시적인 측면에서 접근하거나 미시적인 방법으로 접근하여 관찰 및 증명되었다. 이러한 복잡한 현상, 구조, 상호관계를 증명하기 위하여 일반적으로 모형을 이용하였으며, 복잡한 구조를 규명하기 위해서는 네트워크 이론 등을 이용하였다.

이러한 복잡한 모형을 다루기 위해서는 방대한 자료와 그 자료들의 처리가 병행하여야 하는데 과거에는 거의 불가능하였다. 그래서 과거에는 축소된 모형을 사용하거나 지나치게 단순한 모형을 대상으로 실험하였다. 그러나, 최근 연구에서는 컴퓨터 등의 IT기술들이 접목되면서, 많은 양의 데이터를 활용한 실험과 시뮬레이션도 가능하게 되었다. 즉, 대용량의 데이터를 사용하는 실시간 모형이나 시뮬레이션이 개발되었으며, 프로그램과 pc

---

35) <http://www.santafe.edu/>

의 발달로 과거의 실험모형을 재해석하기도 한다. 또한, 인터넷의 등장으로 수많은 데이터들과 관련 자료들의 접근이 용이해졌으며, 시간 및 공간의 제약이 줄어들어 연구의 접근방법과 형태가 다양해지고 있고 있다. 또한, 컴퓨터의 하드웨어의 발전속도와 소프트웨어의 발전 속도는 상상을 초월하고 있으며, 모형의 특징을 제대로 반영한 설계는 복잡하면서 엄청난 양의 데이터를 실시간으로 처리할 수 있는 단계까지 이르렀다.

따라서, 과거의 단순하고 소규모의 실험모델들이 많은 데이터를 활용한 복잡한 모델로 가능하게 되었으며, 과거에 증명되지 않았거나 불가능한 모델들이 새롭게 재현되고 있다. 그리고 인터넷의 발달로 새로운 모델과 접근방식이 생기기 시작하였다.

현재 우리가 살아가는 세상에서 발생하는 복잡하고 다양한 자연 및 사회학적 현상을 이해하기 위한 다양한 방법들과 학문영역을 넘나들면서 연구는 복잡계의 범위를 넓히는 계기가 되고 있다. 이러한 복잡계를 좀 더 구체적으로 이해하기 위해서는 복잡계가 발생하게 된 근원과 배경지식이 필요하다. 우선 복잡계가 발생하게 된 근원을 살펴보면 다음과 같다.

현재 우리 주변에 존재하는 시스템들은 복잡한 사회, 경제, 문화 등의 다양하고 급격한 변화는 시간의 흐름에 따라 아주 다이내믹하게 작동하고 있다. 이러한 다이내믹한 현상들을 자연과학과 공학분야에서는 최소한의 모형을 선정하고 설명할 수 있는 많은 모델들을 만들고 증명하였으며, 이에 따라 많은 현상들을 이해하고 응용하게 되는 초석이 되었다. 그러나 최근 컴퓨터, 인터넷, 휴대폰 등의 통신망과 전자기기 등의 IT산업의 발전하면서 그와 관련된 인간의 모든 관련 배경들이 기하급수적으로 증가하였으며 상호연관관계는 더욱 복잡해졌다. 이러한 이유로 많은 거시적인 변화들은 현재의 이론체계들로는 더 이상 예측이 곤란한 부분이 생기기 시작하였다. 즉 다이내믹한 변화의 주기들을 예측할 수 없게 되었다. 또한 사회 및 경

제분야 등 인간과 관련된 다이내믹스는 더욱 복잡해지고 있다.

예를 들면, 지금까지의 기상예측은 과거 몇십년의 데이터를 수집하고 분석하여 현재와 미래에 발생하는 기상의 정보를 예측하는 시스템을 구축하여 시행하고 있다. 그러나, 최근의 기상현상들은 과거의 데이터와 전혀 다른 양상을 보이고 있어, 통계학적 기상예측을 신뢰성이 떨어지고 있다. 또한, 슈퍼컴퓨터를 이용한 기사예측도 불과 서너시간만을 예측하는 수준으로 떨어져 기상을 예측하는데는 새로운 뭔가가 필요하게 되었다. 즉, 과거의 데이터를 기준으로한 확률적 모형은 사회현상이나 자연현상들을 예측하기에는 많은 변수가 존재하고 그 변수를 찾아내기는 여간 힘들게 아닐뿐더러, 판단기준을 세우기조차 힘든 상황이다.

따라서, 복잡계는 이러한 수많은 구성요소들의 다이내믹한 현상들이 상호관계에 의해 새로운 현상과 질서를 나타내며 이를 창발현상(Emergent behavior)이라고 한다. 이러한 창발적인 질서를 만들어 가는 과정을 자기조직화(Self-organization)라고 한다. 창발현상은 임계점(Critical point)을 전후하여 급격한 변화가 이루어진다. 이러한 복잡계는 다음과 같이 다양하게 정의할 수 있다.<sup>36)</sup>

---

36) 윤영수, 채승병, 복잡계 개론(삼성경제연구소, 2005)

머레이 겔만(1969, 노벨 물리학상)는 복잡계는 그 특징이 구성요소들을 이해하는 것만으로는 완벽하게 설명되지 않는 시스템이다.

W. 브라이언 아서(산타페)는 복잡계란 무수한 요소가 상호 간섭해서 어떤 패턴을 형성하거나, 예상 외의 성질을 나타내거나, 각 패턴이 각 요소 자체에 되먹임되는 시스템이다. 즉 복잡계는 시간의 흐름에 따라 끊임없이 진화하고 펼쳐지는 과정에 있는 시스템이다.

하버트 A. 사이먼(1978, 노벨 경제학상)은 복잡계란 많은 구성요소들이 그들 사이에 비교적 많은 연관관계를 가져서, 각 구성요소의 행동이 다른 요소들의 행동에 좌우되는 시스템이다.

루스텀 F. 이스마길로프(하버드대 화학과교수)는 복잡계는 그 변화가 초기 조건이나 작은 요동에 매우 민감하거나, 독립적으로 상호작용하는 아주 많은 구성요소를 가지고 있거나, 다양한 진화 경로에 가능성을 갖고 있는 시스템이다.

제롬 L. 싱어(예일대 심리학과 교수)는 복잡계란 상호작용하는 수많은 행위자를 가지고 있어 그들의 행동을 종합적으로 이해해야만 하는 시스템이다. 이렇나 종합적인 행동은 비선형적이어서 개별요소들의 행동을 단순히 합해서는 유도해낼 수 없다.

이러한 다양한 정의에서 창발현상이 일어나기 위한 복잡계의 중요한 특징들을 살펴보면 다음과 같다.

1) 복잡계는 엄청나게 많은 구성요소들이 존재하며, 이들 개체간에 수많은 상호작용들이 일어나고 있다. 즉 복잡계에서는 구성요소간 상호관계를

분석하는 것이 가장 중요한 특징이다.

2) 복잡계에서 구성요소들간의 상호관계를 비선형적이 흐름을 가진다. 이러한 비선형성은 작은 변화에도 구성요소간의 상호관계를 통해 엄청난 파급효과를 가져온다. 즉 임계점을 전후하여 급격한 변화가 일어난다. 이러한 비선형적인 상호작용속에도 발견되는 새로운 질서가 있으며, 이 반복되는 패턴의 질서의 성질을 자기유사성(Self-similarity)이라 하고 이러한 구조를 프랙탈(Fractal)이라고 한다.

3) 복잡계 구성요소들간의 상호관계를 피드백 고리(Feedback loop)를 발생시켜, 전혀 예기치 못한 결과를 초래할 수도 있다.

4) 복잡계는 열린시스템(Open system)에서 쉽게 발생하며, 그 경계면을 불분명하다. 즉, 열린시스템에서는 자유로운 입출입에 의해 변화의 요소들이 극대화 된다.

5) 복잡계는 1회성에 그치지 않고 끊임없이 다른 시스템으로 적응해 나간다. 즉 끊임없이 변화에 적응하는 구성요소들을 복잡적응계(Complex adaptive system)이라 한다.

이러한 복잡계는 구성원들간의 상호관계에 의한 창발현상을 이해하는 것이 중요하며, 어떻게 상호작용들이 창발현상을 만들어 내는가에 대한 지금까지 밝혀진 이론들을 이해하는 것이 중요하다. 이러한 이론들은 20C 중반까지 수학인 분석이 어렵고 너무나 많은 구성요소들로 연구에 한계를 가지고 있었다. 그러나, 20C중반이후 컴퓨터의 등장으로 복잡계 연구가 활성화되면서 많은 이론들이 밝혀지게 되었다(로렌즈의 혼돈이론, 뢰엘과 타켄스의 난류, 만델브로의 프랙탈과 자기유사성, 하겐의 자기조직화 등)

복잡계에서는 복잡한 구성원들간의 상호연관관계를 어떻게 분석하는가에

대한 연구방법들은 일반적으로 이해하기 쉬운 모형을 통해 분석한다. 이 모형을 분석하는 대상이나 방법을 통해 나눈다. 첫째 구분기준은 구성원을 대상으로 분석하는 방법과 시스템을 변화시키는 변수를 대상으로 모형을 만드는 분석하는 방법이 있다. 둘째는 정량적인 방법과 정성적인 방법으로 모형을 만드는 방법이다. 정성적인 방법은 실제 현상을 관찰하여 귀납적 모형을 제시하고 연역적인 방법을 이용해 적절한 가정으로 해석하는 방법이다.

정량적으로 해석할 수 있고 각각의 구성원들에 대한 모형을 다루는 행위 자기빈 모형은 구성원들간의 상호연관관계를 환경변수들에 의해 모형으로 구성하여 창발이 일어나는 모형을 정량적으로 연구하는 방법이다. 또한, 네트워크 모형은 각각의 구성원들간의 상호연결구조와 메커니즘에 대한 연구방법이 있다. 이 네트워크 모형을 이용한 분석은 아직 많은 연구가 필요한 분야이다. 또한 시스템의 변수에 포커스를 맞추고 정량적으로 표현할 수 있는 연구방법으로 시스템 다이내믹스 방법이 있다. 즉 복잡한 연관관계에 의해 어떻게 되먹임 현상이 발생하는지에 대한 모형화 연구방법이다.

정성적인 해석방법으로 각각 구성원들을 대상으로 어떤 목표를 가지고 행동할 때 결과적으로 어떤 변화를 생길 것인가를 해석하는 모형으로 게임 이론과 의사결정이론이 대표적이다. 또한, 정성적인 해석방법으로 변수에 의한 대표적인 해석모형은 비선형시계열 모형 연구방법이다. 이 모형화 방법은 시스템의 변수들이 시간의 변화에 따라 어떻게 변화하는지를 변수들의 함수로써 표현하는 방법이다. 이러한 해석모형으로는 미분방정식 모형, 비선형시계열 등의 연구방법이 있다.



<그림 4> 복잡계의 분류방법

복잡계는 위와 같은 분류와 연구방향으로 지금까지 끊임없이 진화중인 이론이다. 따라서, 복잡계를 전체를 이해하기 위해서는 역사적인 배경에서부터 맥락을 이해해야 된다.<sup>37)38)</sup>

15C 이전의 유럽에서는 플라톤과 아리스토텔레스가 주장한 우주의 중심은 지구이고 달을 경계로 천상의 세계와 지상의 세계가 구분되며, 천상에서의 자연법칙과 지상에서의 자연법칙은 다르다는 세계관이 고대 그리스, 로마 시대에서 중세 유럽에 이르기까지 천년이 넘는 시간 동안 유럽인들의 사고를 지배하고 있었다.

16C 뉴턴의 등장 아리스토텔레스의 자연철학은 16세기에 접어들면서 코페르니쿠스, 케플러, 갈릴레오 갈릴레이 등의 학자들에 의해 의문이 제기

37) 위키피아 <http://www.wikipedia.org>

38) James T. Cushing. (1998), 송진웅 옮김, 물리학의 역사와 철학(Philosophical Concepts in Physics), 북스힐(2006)



되기 시작하였으며, 17세기 영국의 아이작 뉴턴의 등장으로 인해 큰 변화가 나타났다. 뉴턴은 그의 저서 '자연철학의 수학적 원리'를 통해 새로운 세계관을 제시하였는데, '관성'과 '보편중력'의 개념을 통해 우주의 모든 물체 사이의 상호인력을 설명하였다. 이는 기존의 아리스토텔레스의 세계관이 천상계와 지상계를 구분하던 것을 불필요하게 만들었으며, 그로 인해 자연과 우주를 바라보는 인식에 커다란 변화가 나타났다

결정론(Determinism)은 과거의 결과가 미래의 원인이 되며, 이 세상의 모든 사건은 이미 정해진 곳에서 정해진 때에 이루어지게 되어 있었다는 이론이다. 결정론에 따르면 우주에서 일어나는 모든 사건과 운동은 이미 그 전부터 결정되어 있으며, 어떤 법칙에 따라 합리적으로 움직인다. 만유인력의 법칙을 발견한 뉴턴과 라플라스 등은 결정론을 지지했다. 특히, 라플라스는 "우주의 모든 입자의 위치와 속도를 안다면 우주의 미래를 예측할 수 있다"고 주장했으며, 초기 결정론의 모태를 만들었다. 흔히 결정론은 라플라스 주의라고도 한다. 숙명론과 자주 혼동하지만, 결정론은 인과관계로 인하여 필연으로 사건이 일어난다는 점에서 숙명론과는 다르다.

결정론에서 자주 거론되는 것으로 자유 의지가 있는데, 자유 의지란 어떤 이가 선택할 수 있는 대안이 있을 때, 신이나 자연 따위에서 벗어나 행동할 수 있는 의지를 뜻하는 낱말로, 결정론은 참이라는 입장에서도 자유 의지가 있다는 의견과 없다는 의견으로 나뉜다. 자유 의지는 없다는 의견은 심지어 인간의 의지마저 결정되어 있다고 주장하고, 자유의지는 있다는 의견은 인간의 의지가 존재함을 주장한다. 이 문제에 대해 대부분 실존주의 철학자는 모든 상황이 결정되어 있더라도 인간은 영속하고도 자유로운 선택을 할 수 있다고 주장한다.

19세기까지 크게 유행했던 결정론은 20세기에 이르러 양자역학이 생겨나면서 새로운 논쟁이 제시된다. 양자역학이론에 따르면 전자와 같은 미시세

계의 물체들은 현상을 정확히 파악하는 것이 불가능하다. 이는 우주의 모든 사건이 결정되어 있다는 결정론의 핵심에 대한 의문이며, 하이젠베르크, 슈뢰딩거, 아인슈타인 등 20세기 초의 과학자들 사이에서 끊임없이 논의된 문제이다. 이 논쟁의 대표적인 사례로는 '슈뢰딩거의 고양이'를 들 수 있다.

19세기 물리학은 결정론에 들어 맞았기 때문에 크게 유행했다. 그러나 20세기에 접어들면서 양자역학이 생겨나게 되자, 모든 것이 미리 정해져 있었다는 결정론이 그 타당성을 잃어갔다. 따라서, 모든 일은 그 사건이 일어날 확률만이 정해졌다는 확률론적 결정론(確率論的 決定論)이 유행했다. 20세기에 프랙탈과 나비효과, 카오스 이론의 등장과 단순하고 예외적인 선형 운동은 과학적으로 증명할 수 있었으나, 복잡하고 예측하기 힘든 비선형 운동은 과학적으로 알아내는 데 실패한 것이 알려져 결정론은 많은 비판을 받았다. 그러나 오늘날까지 결정론적 개념 규정은 미완이며, 자유의지의 존재도 분명하지 않다.

결정론의 또 다른 논쟁 중 하나는 '미래에 대한 예측'에 있다. 결정론에 따르면 현상을 정확히 파악하면 미래를 예측할 수 있다. 이는 천체역학의 발전을 통해 증명되었으며, 지표에서도 물체의 역학적 운동의 대부분은 수학적 계산을 통해 파악할 수 있다. 그러나 현재의 상황을 분석해도 미래에 대한 예측이 어렵거나 불가능한 사례가 많은데 대표적으로 일기예보, 조류의 흐름, 삼체문제, 심장박동 등이 있다. 이러한 문제들은 20세기에 이르러 프랙탈, 결정론에 입각한 혼돈이론 등을 통해 설명할 수 있는데, 이러한 설명 방식도 선형적인 운동만을 설명할 수 있다는 한계가 있다. 오늘날에도 혼돈이론은 계속 연구되고 있으나, 이것이 과학 전반의 개념적 토대들로서 얼마나 중요한지에 대해서는 여전히 논의되고 있다.

자연에 존재하는 가장 복잡한 구조는 생명체이기 때문에 이러한 생명현상을 설명하기 위해서는 최근 기존의 환원주의적 사고보다는 전체주의적

(Holism)을 중심으로 바라보는 복잡계가 연구되어지고 있다.

환원주의는 복잡한 현상의 원인을 보다 단순한 현상에서 구하는 것으로서 복잡한 구조와 속성이 부분을 통해서 설명될 수 있다는 신념이다. 이러한 환원주의는 기계론적 세계관과 함께 근대과학의 중심적인 특성으로 자리잡았다. 생물학에서 환원주의는 19세기 중반에 기존의 자연철학의 생기설에 대한 비판과 반성으로 나왔다. 급진적인 유물론적 환원론자들은 생명현상을 그 생명을 구성하는 각 부분들의 물리적, 화학적 설명으로 환원되어야 한다고 주장하였다. 이러한 기계적이고 환원적인 접근은 생물에 대한 미시적인 눈을 제공함으로써 생체분자를 규명할 수 있었고 생물의 독립적인 기능단위인 세포를 발견하게 하였다. 세포학의 도입은 생물학을 탄탄한 기초 위에 올려놓았고 생물학 발전의 획기적인 장을 열었다. 그러나 본격적인 생물학의 발전은 20세기 초 슈뢰딩거가 생물학에 양자역학을 도입함으로써 이루어졌다. 양자역학자로서 그의 기계론적 환원주의에의 우려에도 불구하고 그 이후 생명현상을 물리적, 화학적 상호작용으로 환원하는 경향은 분자생물학의 기본 주류가 되었다. 더불어 유전정보의 매개체인 DNA의 발견과 구조결정, 그리고 유전정보전달의 기전 규명으로 생명과학은 혁명적 도약의 순간을 맞이하였다. 이와 같이 근대과학의 환원주의는 생명현상의 이해에 있어서 새로운 돌파구와 실마리를 제공하였다.<sup>39)</sup>

이러한 과거의 과학적 사고의식은 환원주의 결정론은 자연의 복잡성이나 생명현상, 미래예측 등은 비선형적인 복잡한 현상에 대해서는 설명해 주지 못하고 있는 실정이다. 그러나, 비선형적인 복잡한 현상을 설명할 수 있는 방법과 틀이 뚜렷이 존재하지 못하였으므로 이러한 연구들이 발전하는데는 한계가 존재하고 있었다. 그러나, 최근 컴퓨터의 발달과 관련 IT산업의 발달로 비선형적인 복잡한 현상을 해석할 수 있게 되었으며, 기존의 다양한 모형을 재실험할 수 있게 되었다.

39) 박장호, 유전자 결정론(환원주의적 시각)에 대한 비판적 고찰

자연의 복잡성을 연구하는 과학은 1970년대 이후 폭발적으로 성장해 온 비평형계 과학과 카오스이론에 그 뿌리를 두고 있다. 초기에 복잡성의 연구는 고전역학으로 대변되는 결정론적 세계관과 20세기 초에 크게 발전한 양자역학 및 통계역학에 기초한 확률론적 세계관 사이의 내재적 갈등구조를 해결하려는 시도에서 시작되었다.

20C초 헨리 푸인카레는 국왕의 상금이 걸린 삼체문제( 태양, 달, 지구와 같이 세 개의 물체로 이루어진 시스템)에서 처음 카오스 현상을 발견했다. 푸인카레 이후 카오스 연구는 자연과학의 주변적 분야에 머물렀다. 그러나 1963년 에드워드 로렌츠라는 MIT의 기상학자가 기상현상의 한 모형에서 카오스 이면의 질서구조를 발견하였으며, 로렌츠 끌개라고 명명된 이 구조는 기상현상의 비예측성대한 일종의 해결책을 제시하여 카오스 연구의 새로운 연구를 마련하였다. 물리학자인 미첼 파이겐바움(970)은 생태계의 한 단순한 모형에서 발견된 카오스가 ‘보편성’을 가짐을 엄밀하게 증명해 냈다. 이후 카오스에 대한 실험과 이론 연구가 세계적으로 확산되며 카오스 연구는 복잡계의 중요한 연구테마로 자리잡게 되었다.

최근 복잡계에 대한 연구는 수많은 구성요소가 서로 상호작용을 하는 네트워크 또는 시스템에 대한 연구로 확산되었다. 복잡계의 경우 어떻게 개체간 상호 협동에 의하여 집단적 패턴이 창발적으로 생성되는가 하는 것은 시스템과 기능적인 측면에서 매우 중요한 과제이다.

복잡계에 관한 연구분야는 신경소자, 심장박동 리듬, 박테리아 군집, 생태계, 면역계, AIDS역학, 생체대사망, 단백질 접힘, 유전학, 동물의 무리짓기, 신경신호 전달망 등 생태계와 생명현상과 초전도 배열계, 중간보기 양자계, 전하밀도파, 레이저와 플라즈마, 무정질 물질, 화학반응계, 화학신호 전달망, 알갱이 흐름, 기후 변화, 난류, 해안선 변화 등 다양한 물리화학계, 주식시장 등의 경제분야, 인간의 생활패턴, 상호연관관계를 연구하는

사회과학 등에서 연구되고 있다.

이런 점에서는 복잡계 과학은 비선형 동역학, 통계물리, 생리학, 계산 인지과학의 방법론의 융합과 물리학자, 생명과학자, 공학자, 인지과학자의 폭넓은 공동작업이 요구된다. 산타페 연구소는 머레이 겔만(1969년 노벨 물리학상 수상), 필립 앤더슨(1977년 노벨 물리학상), 케네스 애로우(1972년 노벨 경제학상)와 같은 세계적 석학들이 모여 자유롭고 새로운 연구를 지향하는 학제적 연구소를 만들고자 하는 구상에서 출발했다. 현재 산타페 연구소는 계산과학, 진화론, 면역학, 뇌과학, 경제학, 사회학, 과학철학 등 다양한 분야를 포괄하며 이를 복잡계라는 새로운 학제적 시각으로 통합하는 연구를 수행하고 있다. 산타페 연구소의 ‘미션 스테이트먼트’(Mission Statement)는 선도적 연구소답게 복잡계 연구와 같이 학문간, 우수성, 신선도, 촉매성 이 네 가지의 일반적인 속성을 가지는 연구를 장려하고 있으며 복잡계 과학의 촉매 역할을 훌륭하게 수행하고 있다.

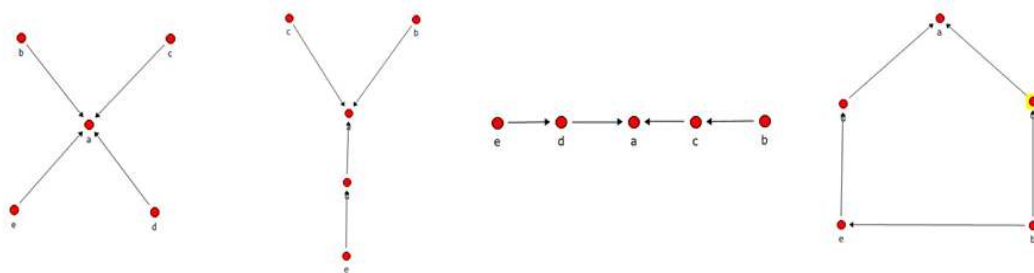
## 제2절 네트워크 분석(Network Analysis)

네트워크와 관련된 이론은 여러 분야에서 발전되어 왔다. 특히, 1930년대부터 사회과학분야에서는 많은 연구가 시작되었으며, 현재까지 가장 많이 응용되는 분야로 사회 네트워크 분석(Social Network Analysis) 또는 사회 연결망 분석이라고 한다. 사회과학에서는 사람의 관계에 대해 과학적으로 접근을 시도한 분야로써, 사람과 사람과의 관계에 대한 가치와 중요성을 중요한 테마로 네트워크이론에 응용하였다.

사회분야의 네트워크에 대한 연구는 사람간의 관계에 대한 속성, 즉 넓은 인맥을 상징하는 마당발의 효과라든지, 누구와 교류를 통해서 업무가 가장 효율적으로 순환되는지, 특정 누구와의 교류를 통해 어떤 가치를 창

조할 수 있는지, 누군의 위치의 중요성은 무엇인지, 어떤 위치가 가장 중요한지, 과연 네트워크상에 누가 가장 이득이 되는지 등 네트워크를 통해 사람의 관계를 분석하는 연구에 응용되었다.

1940년 미국 MIT대의 알렉스 바베라스 교수와 그 제자인 해롤드 리빗의 실험에 의해서 네트워크의 참여자들 사이의 관계에 대한 실험을 실시하였다. 바베라스와 리빗은 그 실험에서 5명이 한 팀을 이루어 스타형, Y형, 체인형, 서클형의 네트워크 유형에서 그들이 공동적으로 해결하는데 소요되는 시간과 효율성을 실험하였다. 이 결과 중심을 가진 네트워크에서 문제의 해결점을 찾는데 가장 빨랐으며 즉각적인 대처 및 실행이 가능하였다. 이로써 네트워크의 중심에 대한 관심을 가지면서 연구가 촉발되었다. 한 네트워크상에서 중심이 가지는 가치와 중요성은 연구가 진행되면서 검증되어졌다. Brass(1984)은 ‘중심에 위치하는 행위자는 집단내의 파워가 강해지기도 한다’, Ibarra(1993)은 ‘중심에 위치하는 행위자는 개인적인 혁신을 용이하게 한다’, Friedkin(1993)은 ‘중심에 위치하는 행위자는 의사결정에 중요한 영향력을 행사할 수 있다’, Baldwin, Bedell and Johnson(1997)은 ‘중심에 위치하는 행위자는 종합적으로 개인적인 성과가 높아진다는’ 등으로 계속적으로 검증되어 응용되고 있으며 중심에 대한 연구결과들이 다른 분야에서도 중요하게 응용되고 있다.

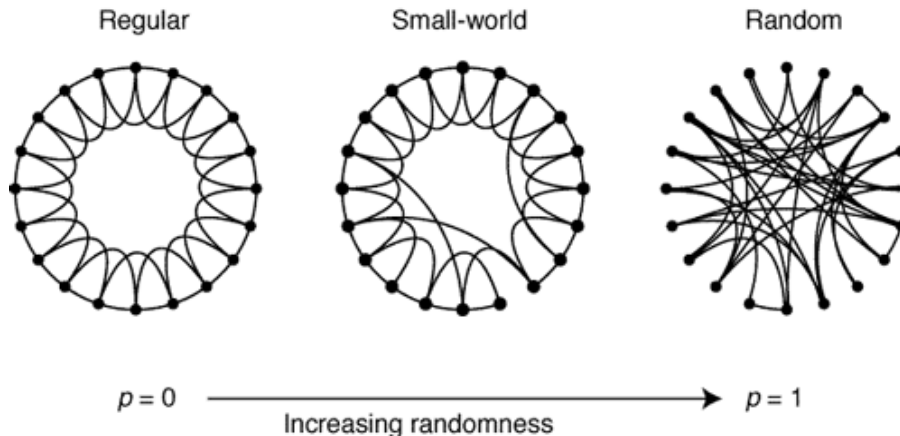


<그림 5> 스타형, Y형, 체인형, 서클형의 네트워크

1960년대 하버드대학의 사회심리학자인 밀그램은 미국 시민들이 평균적으로 여섯 단계(six degrees of separation)만 거치면 서로 연결되어 있음을 증명하였다. 즉, 임의로 선정한 사람이 6단계만 거치면 서로 연결되었다는 사실을 과학적으로 증명하였다. 실험은 네브래스카주의 오마하에 사는 사람을 임의로 추출해서 160통의 편지를 띄우고, 그 편지를 최종적으로 받는 사람은 보스톤에 사는 한 증권 브로커였다. 편지 내용에는 대략 "이 편지는 보스톤 xxx에 사는 증권 브로커에게 전달되어야 할 편지입니다. 이 증권 브로커의 이름을 참조해서, 귀하가 알고 계시는 분 중 가장 이 사람에게 근접하다고 생각되는 사람에게 발송해 주시기 바랍니다."라고 적혀 있었다. 최종적으로 160 통의 편지중에 증권 브로커에게 전달되는데 성공한 편지는 42통이었으며, 전달된 편지가 평균 5.5명에 거쳐 도착한 것으로 나타났다. 이 실험은 실험의 대상과 범위에 대해 많은 문제점들이 제시되어 왔지만, 한편으로는 얼마나 좁은 세상인지 보여주는 스몰 월드 효과(Small World Effect)를 보여주는 시초가 되었다.

1998년 코넬 대학의 Duncan Watts와 Steven Strogatz는 두 명의 학자가 “Nature”지에 기고하여 스몰 월드 네트워크(small-world network)를 발표하였다. 이 논문에서는 일정한 규칙에 따라 인접한 곳과 일정한 숫자로만 링크되는 ‘regular network’와 무작위로 서로 연결되어 있는 ‘random network’의 중간쯤에 해당하는 ‘small-world network’을 발표하였다. 즉, 좁은 세상의 구조에서는 노드간의 지름길(shortcut)이 존재한다는 것이다. 예를 들면, 어떤 집단에 감기 걸린 사람이 한 명 나타났을 때 모든 사람이 주변 몇 사람만 일정한 숫자로 연결되어 있는 경우보다 각 구성원이 무작위로 여기 저기 연결되어 있는 경우 감기 전파 속도가 빠르다. 그런데 이상하게도 엉뚱한 곳으로 연결된 구성원이 딱 몇 사람만 있는 경우에도 감기가 급속하게 퍼지는 것이 관찰되었다. 또는, 우리 신체의 신경망이 망가져서 뇌가 동작을 멈추는 것이 꼭 모든 뇌세포가 다 문제가 생겨

서 그런 것이 아니고 한 두개의 이상한 연결을 가진 셀이 파괴를 순식간에 넓은 범위로 퍼뜨릴 수 있다는 것 등을 스몰 월드로 설명할 수 있다.<sup>40)</sup>



1974년 미국의 사회경제학자인 마크 그레노베터의 ‘직업찾기(job search)’에서는 관계의 정도가 주는 차별성에 대한 연구의 시초가 되었다. 직업을 처음 갖는 사람들이 어떻게 실제 취직에 대한 정보를 얻는가에 대한 조사결과, 실제 그 사람이 직업을 갖게 되는 정보는 약한 연결의 사람들에게 얻은 정보로 인해 취직이 되었다는 결과가 나타났다. 즉 관계의 정도가 약한 사람에 대해 새로운 정보를 얻을 수 있다는 결과를 발표해 관계의 정도에 대한 연구의 뿌리를 두게 되었다.

이와 같이 사회 네트워크 분석에서는 오랜 세월동안 지속적인 연구가 진행되었으며, 현재에서 연구가 진행되고 있다. 사회과학분야에서는 네트워크 이론이 현재까지 연구되어지는 가장 큰 특색은, 기존의 다양한 이론들은 구조의 포괄적이고 추상적인 의미나 구조의 역할이나 제약 등의 구조 자체만을 특성을 분석하였으나 사회 네트워크 분석은 구조의 근본이 되는 행위자들에 대한 관계를 분석하여 구조를 분석하는 구체적으로 접근한다. 특히, 행위자들을 동등한 입장에서 시작하기 때문에 구조속의 행위자들에 대

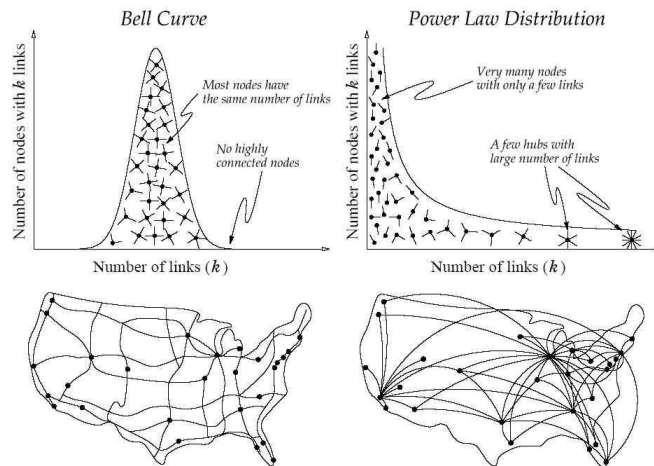
40) 이명현 경영스쿨 [http://www.emh.co.kr/xhtmll/small\\_world\\_effect.html](http://www.emh.co.kr/xhtmll/small_world_effect.html)



한 역할이나 관계가 더욱 중요한 역할을 담당한다. 이는 현재의 인간 개개인의 분석방향과 일치하는 부분이 있다.

또한, 네트워크 이론은 응용수학과 물리학 분야에서도 다루고 있으며, 수학에서는 그래프 이론에서 시작되었다. 네트워크의 수학적 접근은 1959년 헝가리의 수학자 폴 에르되시(P. Erdős)와 알프레드 레니(Alfréd Réney)에 의해 그래프이론(Graph theory)이 도입되면서 현대적인 네트워크의 이론이 정립되기 시작하였다. 에르되시와 레니에 의한 무작위 네트워크는 네트워크의 노드에 의한 연결분포수는 일반적인 종모양의 분포를 따른다고 가정하고 분석하였다. 일반적으로 에르되시와 레니의 네트워크는 무작위로 나타나는 네트워크 현상에 적합한 모델로 인정되어 왔다.

1999년 들어 바바리시(A.L. Barabasi), 알베르(R. Albert), 정하웅에 의해 인터넷의 웹페이지를 설명하는 척도 없는 네트워크(Scale-Free Network) 모형을 제시하게 되었다. 기존의 무작위 네트워크에서는 연결정도 분포가 정규분포를 따르고 것으로 알려져 왔으나, 실제 복잡한 세계에서는 연결정도 분포수는 정규분포보다는 멱함수(Power law) 곡선을 따르는 것으로 증명되었다. 즉 허브의 발견이다. 예를 들면, 물리적으로 연결된 인터넷망에서 일반적으로 노드 각각의 연결수는 평균을 중심으로 극단으로 갈수록 최저 및 최고 연결수가 거의 없는 정규분포를 따를 것으로 예상되었으나, 실제 노드의 연결수 분포를 확인해 본 결과 최고와 최저의 연결수가 상당히 존재한다는 사실이 밝혀지고 있다. 이러한 사실로 인하여 멱함수를 따르는 네트워크가 상당히 많이 존재하다는 사실이 밝혀졌다. 이를 ‘척도 없는 네트워크(Scale-Free Network)’이라고 명명하였다.



<그림 7> Random Network and Scale-Free Network

이러한 네트워크 이론들의 연구가 급속하게 발전하게 된 근본적인 이유는 컴퓨터와 인터넷 등의 IT기술들의 발달로 과거에 수행되지 못한 모형들이나 거대한 네트워크의 분석, 데이터의 수집 및 처리가 가능하게 되었다. 이러한 대용량의 데이터를 처리할 수 있는 능력들은 네트워크의 놀랄만한 성과를 이루어 내고 있다.

1946년은 최초의 컴퓨터 에니악이 개발되면서 컴퓨터가 등장하였다. 1954년 실리콘 트랜지스터의 등장으로 1955년부터 트랜지스터 컴퓨터가 시작되었다. 기존의 에니악으로 대표되는 1세대는 진공관을 사용하였다. 1957년 IMB의 백커스에 의해 최초의 진보된 프로그램 언어인 포트란이 개발되면서 본격적으로 프로그램이 개발되기 시작하였다. 1958년 텍사스 인스트루먼트 사의 킬비가 하나의 반도체 기판 위에 저항, 축전기, 트랜지스터를 모두 올려두고 그 기판에 부착된 선으로 연결시킨 IC회로를 이용한 컴퓨터를 만들면서 소형화와 가격 하락을 가져왔으며, 마이크로프로세서 시초가 되었다. 1960년 코볼개발, 1964년 베이직의 개발, 1968년 마우스의 개발로 유저 인터페이스 개념이 도입되기 시작하였다.

1964년 인터넷의 시초인 ARPANET 프로젝트는 미국 각지에 산재된 여러 상표의 컴퓨터들이 서로 대화할 수 있도록 패킷 교환 이론을 이용한 네트워크를 구축하는 것을 목적으로 개발하였으며, 군사적 목적에 의해 개발되었다. 1970년 유닉스의 개발과 1971년 C언어의 개발로 현재의 운영체제의 기본과 프로그램 언어의 기초가 완성되므로 엄청난 계산이 가능하게 되었다. 또한, 1976년 최초의 슈퍼컴퓨터의 크레이의 개발로 대단위 과학계산이 가능하게 되었다. 1977년 마이크로소프트의 개인 운영체제 DOS의 등장과 1978년 인텔의 8086의 등장으로 개인도 컴퓨터를 사용할 수 있는 기회를 제공하게 되었다. 1980년대 이후 수많은 OS와 프로그램언어, 하드웨어 등의 발전으로 인류역사상 새로운 전기를 마련하게 되었다. 이에 모든 인류에게 영향을 미치면서 과학기술이 엄청나게 발전하게 되는 계기가 되었다. 현재에는 IT분야라고 불리우며 엄청난 과학기술의 개발이 진행되고 있다.

현재 월드와이드웹, 인터넷, 단백질 상호작용 네트워크, 사회 네트워크 등 많은 분야에서 복잡계 네트워크 이론이 적용된다는 연구가 발표되면서 활발히 연구가 진행되고 있다.

### 제3절 네트워크 분석(Network Analysis) 방법

네트워크 분석에서는 근본적으로 네트워크에서 노드간의 연관관계에 따른 노드의 역할이 네트워크에 어떤 역할을 담당하거나 어떤 영향을 미치는 것을 파악하는 것이 중요하다. 일반적으로 네트워크의 중심을 분석하기 위하여 중심성 분석<sup>41)</sup>(Centrality Analysis)을 실행하며, 이 분석은 노드가 중심에 어느정도 근접한가를 표현하는 방법이다.

---

41) G. Sabidussi, The centrality index of a graph. Psychometrika, 31(4)(1966) 581-603

사회과학분야에서는 네트워크의 중심에 대한 중요성은 1940년 미국 MIT의 알렉스 바베라스(Alex Bavelas)와 해롤드 리빗(Harold Leavitt)의 실험에서 스타형과 Y형의 스타일을 가진 네트워크가 문제해결과 즉각적 시행적인 측면에서 우수하다는 연구결과가 발표되면서, 네트워크의 중심분석에 대한 연구를 촉발시키는 계기가 되었다. 1984년 브래스(Brass)는 중심에 위치하는 행위자는 집난내 파워가 강하다고 발표하였으며, 1993년 아이바라(Ibarra)는 중심이 개인적인 혁신이 용이하기 이루어진다고 발표하였으며, 같은해 프라이드킨(Friedkin)은 중심이 의사결정에 중요한 영향력을 미친다고 발표하였다. 따라서, 네트워크에서 노드의 중심을 어떻게 분석하는 것이 중요한 이슈가 되고 있으며, 본 연구에서는 중심성 분석(Centrality Analysis)을 통해 네트워크의 중요성을 분석하였다.

네트워크의 중심성 분석방법은 각각의 행위자를 표현하는 노드와 상호연관관계를 의미하는 링크에 네트워크를 구성하게 되어있으며, 이들의 상호관계를 규명하는 것이 중요하다. 특히, 어떤 노드가 허브역할을 하는지, 어떤 노드가 중계자의 역할을 하는지 등의 중심에 대한 해석이 가장 중요하며, 전체 네트워크에서 각각의 노드들에 대한 역할, 위치, 특성, 영향력을 파악하는 것이 중요하다. 또한, 네트워크내에서 노드들의 상대적 위치나 절대적 위치를 파악하는데 유용하게 활용할 수 있다.

일반적으로 중심성을 분석하는 방법은 한 점에 연결된 다른 점들의 수를 측정하는 연결정도중심성(Degree Centrality), 한 점이 다른 노드들간의 네트워크를 구축하는데 중계자 혹은 브릿지 역할을 측정하는 매개중심성(Betweenness Centrality), 한 점에서 다른 점에 얼마만큼 가깝게 있는가(distance)를 측정하는 근접중심성(Closeness Centrality)이 있다.

연결정도중심성(Degree Centrality)은 한 노드에 이웃한 모든 노드들과의 링크를 수를 측정하여 하나의 노드에 얼마나 많은 링크가 연결되어 있

는가를 나타내는 지표로써, 허브역할을 담당하는 노드를 찾는 방법으로 활용되고 있다.

$$\text{상대적 연결정도} = \frac{\text{연결정도}}{(\text{네트워크내 전체 점의 수} - 1)}$$

매개중심성(Betweenness Centrality)은 네트워크 내에 한 노드가 매개자나 중계자의 역할을 담당하는 노드를 찾는 방법으로, 연결정도가 많고 적음에 상관없이 두 네트워크간 혹은 그룹간의 중재역할을 하는 노드를 찾는 방법으로 정보유통이나 소통역할을 담당한다.

$$\text{상대적 매개중심성} = \frac{\text{두 노드사이에 경유하는 노드의 횟수}}{(g-1)(g-2)/2}$$

※  $(g-1)(g-2)/2$  : 최대 가능한 매개중심성 횟수

※ 범위:  $0 \leq \text{매개중심성} \leq 1$

근접중심성(Closeness Centrality)은 네트워크에서 각 노드간의 거리의 개념을 이용하여 최단거리의 합을 이용하여 전체 네트워크에서 가장 중심이 되는 노드를 찾는 방식으로, 모든 노드로부터 가장 짧은 거리에 있는 노드이다(전체 네트워크에서 정보, 권력, 영향력, 사회적 지휘를 확보와 접근이 쉬운 노드를 찾을 수 있음)

$$\text{상대적 인접중심성} = \frac{(\text{노드수} - 1)}{(\text{두노드간 거리의 합})}$$

※ 범위:  $0 \leq \text{근접중심성} \leq 1$

대규모 네트워크 구조 분석시 노드와 링크가 너무 많아 노드와 링크에

대한 분석에 많은 어려움이 있으며 시각적으로 표현하는데 한계가 있어, 전체 네트워크에서 중요한 역할을 담당하는 링크만을 남겨두고 나머지를 제거하는 방법으로 네트워크의 간략화 및 최적화하는 방법을 사용한다. 일반적으로, 방향성이 없고 모든 링크가 연결된 네트워크에서의 최적화 방법은 다양하나, 일반적으로 알려진 방법들은 절대기준방식, 최근접이웃 방식(Nearest Neighbor Graph : NNG), 최소비용 신장트리(Minimum Spanning Tree : MST), 패스파인더방식(Pathfinder : PFNet)이 있으며, 현재까지 패스파인더방식이 최적화 알고리즘으로 효율성이 좋은 것으로 알려져 있다.<sup>42)</sup>

패스파인더방식<sup>43)</sup>(Pathfinder : PFNet)은 방향성이 없고, 모든 링크가 연결된 가중치를 가지고 있는 네트워크에서 삼각부등식(Triangle inequality)을 위반하는 경로를 제거하는 방식으로 Schvaneveldt, R. W. (Ed.) (1990)에 처음 개발되었으며, 본 연구에서 패스파인더 방식을 사용하였다. 사용된 패스파인더방식(Pathfinder : PFNet)<sup>44)</sup>은 두 개의 파라미터( $r=\infty$ ,  $q=n-1$ )에 의해 결정되며, 파라미터  $r \in [1, \infty]$ 에서는 Minkowski 거리공식에 의해 방향성이 없는 두 노드간에 연결된 거리를 고려하는 방법으로, 아래의 식에 따라 가중치가  $r=1$ 이면 링크 가중치의 합이 경로거리가 되고,  $r=\infty$ 가 될수록 전체 링크에서 가중치가 가장 최대인 값이 경로거리가 된다.

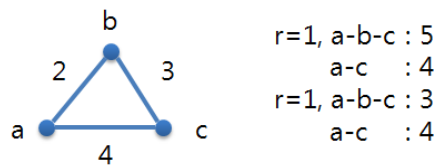
$$D = (\sum_i d_i^r)^{1/r}$$

42) Jae-Yun Lee, A Study on the Network Generation Methods for Examining the Intellectual Structure of Knowledge Domains, Korean Society for Library and Information Science 40(2006) 333-355

43) R. W. Schvaneveldt, F. T. Durso, and D. W. Dearholt, Network structures in proximity data. In G. Bower (Ed.), The psychology of learning and motivation: Advances in research and theory, Vol. 24(Academic Press, New York, 1989)

44) Quirin, A; Cordon, O; Santamaria, J; Vargas-Quesada, B; Moya-Anegón, F, A new variant of the Pathfinder algorithm to generate large visual science maps in cubic time, Information Processing and Management, 44, p.1611-1623(2008)

파라미터  $q \in [2, n-1]$ 에서는 삼각부등식(Triangel inequality)을 사용하기 위한 노드 사이의 경로를 산출하는 최대 링크수를 나타내면,  $q$ 는 2부터 설정 할 수 있다( $n$  : 전체노드수,  $q=n-1$ ) 예를 들면, 세점  $a, b, c$ 에서  $a$ 에서  $c$ 로 가는 경로에서  $r=1$ 이면  $a-b-c$ 의 경 경로 가중치는 5이며  $a-c$  경로 가중치는 4이므로 삼각부등식이 성립되어 제거되지 않지만,  $r=\infty$ 에서는 Minkowski 거리공식에 의해  $a-b-c$  경로 가중치는 3이고,  $a-c$  경로 가중치는 4이므로  $a-c$  경로는 제거된다(그림3 참조)



<그림 8> 패스파인더 예시

<표 2> 네트워크 최적화 방법의 개념 및 장단점 비교

구분	절대기준 방식	최근접이웃 방식 (Nearest Neighbor Graph)	최소비용 신장트리(Minimum Spanning Tree : MST)	패스파인더방식 (Pathfinder : FFNet)
기준	· 절대적 가중치 값	· 상대적 가중치 기준(노드별) · 각 노드별 기준 갯수(k)에 의한 가중치 우선순위	· 상대적 가중치 기준(전체) · 전체 네트워크에서 링크 가중치의 합이 최소가 되는 트리	· 상대적 가중치 기준(전체) · 각부등식(Tringle inequality)를 위반하는 링크를 제거하는 방식
개념	· 기준 가중치 값 이상의 링크만 표시하는 방법 · 전체에 일정한 가중치기준을 적용	· 각 노드에 연결된 링크의 갯수(k)를 기준으로, 그 개수에 따라 가중치의 높고 낮음의 우선순위에 따라 표시하는 방법	· 최소한의 가중치로 전체가 연결되도록 링크 가중치가 높더라도 전체 구조상 중복된 링크 제외 · Kruskal(1956)과 Prim(1957)의 알고리즘가장 널리 알려져 있음	· MST방식은 같은 가중치를 가지는 링크일 경우 우선순위의 기준이 없으나, PFNet의 경우 우선순위가 있기 때문에 더욱 상세한 네트워크 표현이 가능함

구분	절대기준 방식	최근접이웃 방식 (Nearest Neighbor Graph)	최소비용 신장트리 (Minimum Spanning Tree : MST)	패스파인더방식 (Pathfinder : FFNet)
장점	<ul style="list-style-type: none"> <li>알고리즘이 단순</li> <li>전체구조의 파악 용이</li> </ul>	<ul style="list-style-type: none"> <li>알고리즘이 단순하며, 단독 고립노드가 나타나지 않음. 군집이 생성되며 세부 구조파악 용이</li> </ul>	<ul style="list-style-type: none"> <li>전체구조의 파악 용이</li> <li>단독 노드가 발생하지 않음</li> </ul>	<ul style="list-style-type: none"> <li>MST보다 디테일한 네트워크 작성 가능</li> <li>단독 노드가 발생하지 않음</li> </ul>
단점	<ul style="list-style-type: none"> <li>시작 노드나 가중치의 기준값의 없음</li> <li>고립노드가 발생</li> </ul>	<ul style="list-style-type: none"> <li>전체 구조의 파악이 어려움</li> </ul>	<ul style="list-style-type: none"> <li>링크의 가중치가 동일할 경우, 일부세부구조가 나타나지 않음</li> </ul>	<ul style="list-style-type: none"> <li>알고리즘이 복잡함</li> </ul>
전체 노드 연결 여부	<ul style="list-style-type: none"> <li>가중치 기준값이 높을 경우, 고립점이 생김</li> <li>가중치 기준값이 낮을 경우, 많은 연결링크</li> </ul>	<ul style="list-style-type: none"> <li>링크의 개수(k)가 1에 가까울수록 서브 그래프, k가 클수록 전체 네트워크는 복잡하고 연결</li> </ul>	<ul style="list-style-type: none"> <li>전체가 연결되어 있음.</li> </ul>	<ul style="list-style-type: none"> <li>전체가 연결되어 있음.</li> </ul>
응용 분야	<ul style="list-style-type: none"> <li>문헌동시인용분석(Small, 1973)</li> <li>공저자분석(Newman 2001) 이외 다수</li> <li>동시출현단어분석(Co-work analysis)에서도 사용</li> </ul>	<ul style="list-style-type: none"> <li>사회과학분야에서 설문, 관찰을 통해 주로 사용하였음(사회연결망 조사)</li> <li>동시인용자료와 계량서지 자료에 적용한 경우는 없음</li> </ul>	<ul style="list-style-type: none"> <li>일반적으로 정보시각화에서 주로 사용</li> <li>계량서지학분야에서는 Noel(2002), Chen과 Morris(2003)정도만 사용되어짐.</li> </ul>	<ul style="list-style-type: none"> <li>인지심리학에서 개발되어 인지적인 연상구조나 연결구조 적당함(검색(검색, 브라우저 UI등)</li> <li>동시분류분석(McCain, 1995) - 생물공학특허분야 분석, 저자동시인용분석(Chen, 2001), 저널동시인용분석(Marion&amp;McCain, 2001), 문헌동시인용분석(Chen, 2001)</li> </ul>



#### 제4절 검색엔진을 이용한 분석방법

인터넷이 급속히 증가하면서 인터넷에는 수많은 데이터들이 넘쳐나고 있으며, 1999년에는 약 3천만개의 웹사이트가 있는 것으로 조사되었으며, 2007년도에는 약 1억개, 현재에도 기하급수적으로 증가하고 있다. 이렇게 기하급수적으로 증가한 인터넷의 실시간 데이터나 중요한 정보를 찾기 위하여 우리는 많은 시간과 노력을 필요로 하게 되었다. 이로 인해, 인터넷 데이터를 활용하여 우리가 필요로 하는 정보를 어떻게 도출할 것인가에 대한 많은 연구가 진행되고 있으며, 텍스트마이닝, 검색엔진, 로봇 등의 다양한 기법들이 도입되고 있으나, 아직까지 많은 연구가 진행되고 있는 실정이다.

이러한 인터넷을 이용한 방법은 여러 가지 중요한 시사점을 나타내고 있다. 첫째, 인터넷에서는 가공되지 않는 무한한 정보가 존재하며, 1차적으로 어떻게 필요한 정보를 얻을 수 있는가에 대한 근본적인 방법 모색(검색엔진을 활용한 방법, 인터넷에 존재하는 데이터를 중심으로 텍스트마이닝을 이용하는 방법 등), 둘째, 인터넷의 정보를 활용하여 직접적인 정보가 아니더라도 간접적인 정보로써 활용이 가능하다는 의미(선거 출구조사 대신 같은 기간의 인터넷에 검색결과에 의한 출구조사, 관련 검색단어의 검색빈도수를 이용한 독감발생 추이 방법 등), 셋째 인터넷을 이용하여 트렌드 및 추세분석이 가능하다는 점(현재 소셜 네트워크(Social Network), 블로그 등을 이용한 개인과 단체의 성향과 트렌드 등을 예측하는 방법 등)

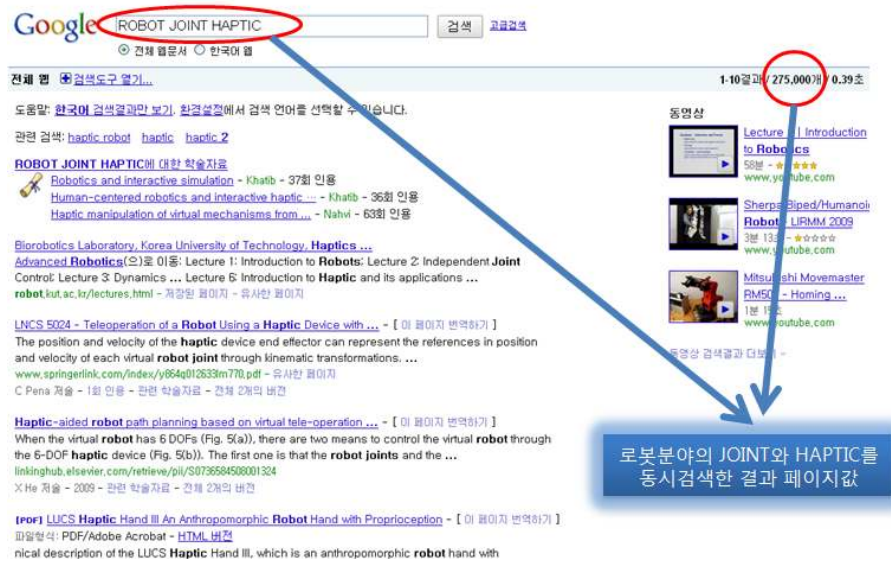
더욱이, 네트워크분야에서도 IT기술과 인터넷을 활용한 새로운 방법들이 도입되기 시작하였으며, 일부 연구에서는 인터넷의 방대한 데이터를 활용하여 데이터마이닝, 텍스트마이닝, 검색엔진을 이용한 방법 등이 연구되고 있다. 예를 들면, 2008년 구글에서는 미국의 109대 상원위원의 선거에서

기존의 출구조사를 한 결과와 구글의 검색결과로 분석한 결과가 유사한 것으로 알려지므로 인터넷의 유용성에 대한 연구가 활발하게 진행되고 있다. 또한, 최근 연구[8]에서는 미국 시민들의 검색단어 빈도수를 이용하여 지역별 독감발생 추이를 분석한 결과, 미국 CDC(The U.S. Centers for Disease Control and Prevention)에서 발표한 실제 독감발생 추이와 유사하고 이런 분석방법을 통해 CDC 발표시기보다 약 2주정도 빠른 결과를 도출할 수 있다[9].

본 연구에서도 인터넷의 다양한 정보(웹페이지, 논문검색, 특허검색, 블로그 등)를 통해 네트워크의 연결고리를 분석하여 가상 네트워크를 도출하고 도출되어진 네트워크를 통해 분석하는 모델을 제시하였다. 특히, 노드 간의 상호 연관관계에 대한 가중치를 인터넷의 구글 검색결과값으로 대체하고 이를 네트워크로 분석하였다. 여기에 사용되어진 구글 검색엔진은 페이지랭크(Pagerank)라는 알고리즘을 통해 구현되었으며, 웹페이지에 연결된 하이퍼링크를 수에 따른 가중치를 분석하여 검색결과값으로 표현한 일종의 네트워크 알고리즘으로 만들어진 검색엔진이다.<sup>45)</sup> 이렇게 구현된 검색엔진은 기존의 검색엔진과 달리 임의로 검색결과를 조작하기 어려워 검색결과에 대한 신뢰성이 높은 것이 특징이다. 최근 구글에서는 Open API(Application Programming Interface)를 제공하여 논문검색사이트, 특허검색사이트, 구글사전 등을 제공하여 다양한 목적을 위해 사용하게끔 공개되어 있다. 일반적으로 API(Application Programming Interface)는 운영체제나 언어가 어떤 기능을 제어할 수 있도록 제공되는 인터페이스였으나, 웹 2.0에서는 웹의 특정한 서비스를 이용하도록 제공되는 인터페이스로 개념이 확장되고 있는 개념이다.

---

45) <http://www.wikipedia.org>



<그림 9> 구글 검색엔진의 활용방법

## 제5절 텍스트마이닝

오늘날 인터넷이 등장하면서 인터넷에서는 수많은 정보들이 넘쳐나기 시작하였다. 이러한 정보들은 지금도 기하급수적으로 팽창하고 있으며 도대체 어디까지 증가할지 정확한 예측이 불가능하다.

그러나, 기하급수적으로 늘어나는 데이터들은 중요한 정보와 중요하지 않는 정보를 구분할 수 있는 한계를 벗어나고 있으며 이를 해결하기 위한 많은 연구들이 행해지고 있으며 활용되고 있다.

오늘날 이러한 정보를 크게 2가지의 형태로 구분하면 정형화된 데이터와 비정형화된 데이터로 구분할 수 있다. 정형화된 데이터의 내용은 기존의 데이터를 활용하기 위하여 일정한 형식과 조건을 만족하는 자료로 가공하여 DB화한 정보이다. 이러한 정보는 전체의 약 20%정도가 자료의 생성, 저장, 재사용하는 정보로 구성되어 있다. 정형화데이터(Structured Data)

의 정보를 추출하고 가공하는 방법을 데이터마이닝(Data Mining)이라고 불리며, 현재 우리가 가장 많이 활용하는 데이터베이스 시스템과 정보분류 체계에 응용되고 있다.

그러나, 대부분의 정보들은 워드프로세서, e-mail, 프리젠테이션, 스프레드시트, PDF 와 같은 복합문서와 인터넷 페이지 등의 비정형 텍스트 형태로 구성되어 있으며, 이를 비정형화된 데이터라고 불리며 문서전체의 80%정도를 차지하고 있다. 텍스트 마이닝 (Text Mining)은 이러한 비정형 정보를 어떻게 활용하는 가에 대한 방법을 말하며, 현재 다양한 분야에서 활용되고는 있으나 아직까지 많은 연구가 필요하다.

텍스트 마이닝 (Text Mining)은 대용량의 데이터에서 사용자가 관심을 가지는 정보를 키워드의 수준이 아니라 Context의 수준의 의미를 찾아내는 프로세스를 의미한다. 즉, 정보의 폭발적인 증가로 많은 부분을 자동적으로 처리할 수 있도록 방법이 필요하게 되었으며, 대용량의 데이터 속에서 숨겨진 패턴을 발견하고 특정 주제와 연관된 데이터를 검색하는 방법으로 발전하고 있다.

텍스트마이닝으로 인해 과거에 생각할 수 없었던 기술실현 방법들은 예상할 수 있다. 예를 들면, 다양한 자료와 시간 등으로 구성된 범죄 기록들 속에 현재 발생한 유산한 형태의 범죄유형을 찾아냄으로써 범죄자나 테러범을 색출, 웹게시판에 올라오는 다양하고 비정형화 되어 있는 수천만건의 고객불만 사항을 특정 카테고리별로 분류하거나 특정문제를 찾아내는 방법, 수많은 환자의 처방내역서에서 당뇨병에 효과적인 치료 패턴을 자동으로 찾아내는 등 다양한 방향으로 응용이 가능하다.

이러한 텍스트 마이닝이 좀더 고차원적인 방법으로 응용되기 위해서는 정보 검색 (Information Retrieval), 기계 학습 (Machine Learning), 통

계학, 자연언어처리(Natural Language Processing), 데이터 마이닝 (Data Mining) 등 여러 학문 분야의 연구 성과들에 기반을 둔 응용분야와 융합연구가 필요하다. 그러나, 융합연구도 최근에 이루어지고 있기 때문에 많은 연구가 필요한 분야이다.

현재까지 텍스트 마이닝은 크게 인터넷분야와 일반적인 데이터를 마이닝 하는 분야로 응용되고 있다. 인터넷을 활용한 데이터마이닝 기법은 인터넷 검색엔진 등에 활용되고 있다.

텍스트마이닝의 일반적인 프로세스는 여러 가지로 알려져 있으나 일반적으로 4단계의 프로세스 거친다. 텍스트마이닝 과정은 [비정형 정보수집 → 정보처리 → 정보추출 → 정보분석] 등의 일반적인 절차를 따르고 있으며, 정보추출과정에서 수학적 모델이나 알고리즘을 통해 유용한 정보를 도출하는 방법이다. 이를 어떻게 활용할 것인가에 대해 검색엔진에 활용하거나 다른 중요한 키워들을 도출하는 등에 활용하고 있다.

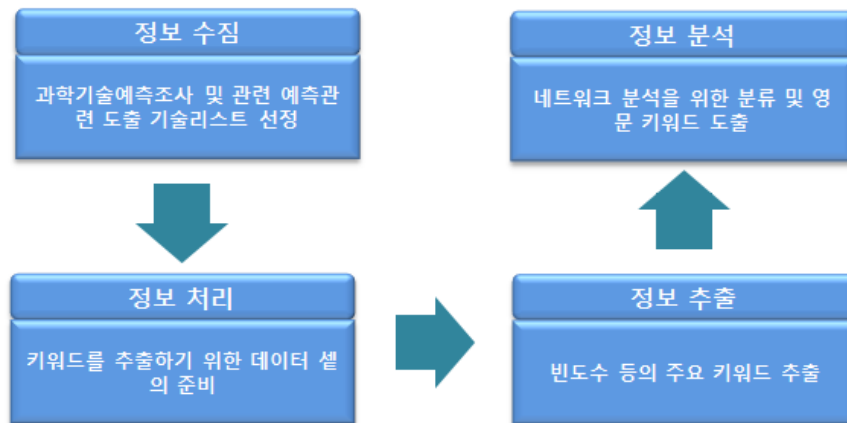
예를 들면, 과학기술에 대한 중요한 키워드들을 분석하는 프로세스에 적용하면 다음과 같다.

첫째, 비정형 정보수집 과정에서는 과학기술예측조사와 관련 분야의 비정형 데이터, 기술자료, 논문 등의 자료를 수집하는 단계로 과학기술예측과 관련된 주요 키워드, 개요서 등의 자료 수집한다.

둘째, 정보처리 과정에서는 도출된 대상 자료와 기술리스트를 중심으로 각 기술에 해당되는 정보를 추출할 수 있는 데이터 소스를 가공하는 단계로 수집된 자료(문서)를 기본으로 관련 키워드의 추출한다.

셋째, 정보추출 과정에서는 수학적 모델이나 알고리즘을 활용하여 필요한 키워들을 추출하는 단계로 가장 핵심적인 분야이다.

넷째, 정보분석단계에서는 최종 키워드들을 우선순위 도출 및 주요 키워드의 분류하는 단계이다.



<그림 10> 텍스트마이닝 프로세스

## 제6절 텍스트마이닝의 정보추출 방법

텍스트마이닝을 위한 정보추출 방법에는 다양한 목적, 조건, 환경 등으로 정보의 추출방법이 다양하며, 정보추출방법은 텍스트마이닝에서 가장 중요한 부분중에 하나이다. 특히, 정보추출 방법에는 수많은 수학적알고리즘과 방법들이 존재하며, 그중 간단하면서 가장 강력한 방법으로는 TF-IDF (Term Frequency - Inverse Document Frequency) 방식을 많이 사용하고 있다.

처음 TF-IDF은 Spark(1972)는 여러 문서에 동시에 출현하는 단어는 범용적인 확률이 높다는 전제아래 역문헌 빈도수를 계산(IDF : Invert Document Frequency)을 제시하였다.

$$idf_i = \log \frac{N}{N_j}$$

$N$ : 전체문헌수

$N_j$ : 용어  $K$ 가 출현한 문헌수

Salton(1976)에 의해 한 문서내에 자주 출현하는 단어는 그 문서를 대표할 수 있다는 명제를 통해 문헌내의 단어빈도를 계산(TF : Term Frequency)하는 방식이 제시되었다.

Wu & Salton(1981)은 이러한 두가지 방식을 Term Weight(가중치 중요도)을 발표하였다.

$$w_{ij} = tf_{ij} \cdot idf_i$$

$$= \frac{tf_{ij}}{\max(tf_{ij})} \cdot \frac{idf_j}{\max(idf_j)}$$

$\max(tf_{ij})$ : 특정문서내 가장 높은 출현빈도를 갖는 단어빈도

$\max(idf_j)$ : 전체문서내 가장 높은 출현빈도를 갖는 문헌수

Buckley & Salton(1988)에 의해 Term Weight(가중치 중요도)의 보정하는 공식을 발표하였다.

$$w_{ij} = (0.5 + \frac{0.5 \cdot tf_{ij}}{\max(tf_{ij})}) \cdot idf_j$$

TF-IDF(Term Frequency - Inverse Document Frequency)에 대해 좀더 자세히 살펴보면, TF-IDF정보 검색과 텍스트 마이닝에서 이용하는 키워드의 가중치를 구하는 방법으로 여러 문서로 이루어진 문서군이 있을 때

어떤 단어가 특정 문서 내에서 얼마나 중요한 것인지를 나타내는 통계적 수치이다. 문서의 핵심어를 추출하거나, 검색 엔진에서 검색 결과의 순위를 결정하거나, 문서들 사이의 비슷한 정도를 구하는 등의 용도로 사용할 수 있다.

TF(단어 빈도수, term frequency)는 특정한 단어가 어떤 범위내의 문서에서 얼마나 자주 등장하는지를 나타내는 단어빈도수 값으로, 특정단어의 빈도수가 높으면 문서내에서 중요하다고 생각할 수 있다(TF값이 높아짐).

DF(문서 빈도수, document frequency)는 특정한 단어가 일정한 범위의 문서들간의 자주 사용되는 지수를 나타낸다. 일반적으로 특정한 단어가 여러 문서에서 빈도수가 높은 경우, 그 단어가 흔한 단어로써 보편적인 단어라는 의미를 나타낸다.

IDF(inverse document frequency)는 DF(문서 빈도수, document frequency)의 역수를 말한다. 특히, 문서간의 역수를 취함으로써 DF가 커질수록 IDF는 감소하고 IDF를 작을수록 IDF를 증가한다.

TF-IDF는 TF와 IDF를 곱한 값을 중요도지수라고 칭하며, 이를 통해 문서의 중요도를 생성할 수 있다.

$$TF-IDF = TF \times \frac{1}{DF}$$

*TF*: 문서내 특정단어의 빈도수

*DF*: 여러문서내의 특정단어 빈도수

*IDF*: *DF*의 역수



예를 들면, 특정 범위내에서 모든 단어들의 빈도수를 구하고, 또한 문서마다의 단어들의 빈도수를 구한후 역수를 취해 곱하며 문서의 중요도가 나타났다.

이러한 단어의 중요도를 응용하여 특정단어에 대한 중요도를 추출하는 정보검색에서도 가장 기본적인 알고리즘으로 널리 사용하고 있다. 즉, 특정단어의 순위 결정은 질의어와 질의어가 포함된 웹 문서들 간의 가중치 및 유사도를 계산하여 높은 값을 가진 가중치를 정렬함으로써 순위가 결정되는 방식을 사용하고 있다. 또한, 대부분의 검색엔진은 이 알고리즘과 다른 알고리즘(벡터방식 등) 등을 혼합하여 사용한다.

그러나, TF-IDF 모델은 문서 내에서 정보검색, 색인과 유사도를 계산하고 검색 순위를 결정할 수 있는 기능을 가지고 있는 반면, 문서내 단어 개수와 문서가 많아지고 문서내에 포함된 단어의 수가 많을수록 검색의 효율성은 떨어지고 방대한 계산으로 너무 많은 시간이 소요되는 단점이 발생하고 있다. 또한, 많은 내용으로 인해 영어단어의 키워드가 정확히 매치되지 못하는 단점이 있으며, 특히 교착어인 한글의 특성상 형태소를 통해 단어를 추출해야 되므로 영어보다 더 많은 시간과 소요되고 있다.

또한, TF-IDF 모델은 최신 트렌드를 반영하는데 많은 어려움이 따르고 있다. 즉 최신 트렌드를 반영한 문서를 수집하는 범위를 확정하는 것이 어려우며, 최신 트렌드를 반영하는 신조어 등장하는 경우, 신기술의 문서일 경우 특정 단어가 포함된 문서가 너무 많아 오히려 중요도가 낮아지는 현상까지 나타나고 있다.

따라서, 미래를 예측하기 위한 트렌드 분석이나 현재 동향을 분석하기 위해서는 위와 같은 문제점을 보완하는 적절한 알고리즘이 필요하다.

구체적인 예를 들어, TF-IDF<sup>46)</sup>(Term Frequency - Inverse Document Frequency)의 소개하면 다음과 같다.

일정범위의 문서군에서 '기후변화(Climature)'라는 단어에 대한 중요도 지수를 계산하기 위해서, 우선 한 문서내에서 100개의 단어속에 '기후변화(Climature)'라는 단어는 10번의 빈도가 된다고 가정하자.

그럼, 한 문서내의 상대적인 TF값은 0.1이 된다. 그런 다음 1,000개의 문서중에서 '기후변화(Climature)'라는 단어가 포함된 문서가 상대적인 DF값은  $200/1000=0.2$ 가 되고 IDF는  $1/0.2=5$ 가 된다. 따라서 TF-IDF는  $0.5(=0.1 \times 5)$ 가 되는 것이다. 실제로는 각 문서마다 TF값이 존재하기 때문에 더욱 복잡하다.

climate 단어	TF 계산		IDF 계산
단어 빈도수	10	문서갯수	200
총 단어수	100	총 문서갯수	1000
TF	0.1	DF	0.2
		IDF	1/0.2
TF-IDF	TF×IDF=0.1×(1/0.2)		

최근, 텍스트마이닝에서 TF-IDF와 같이 문헌과 용어에 대한 유사도와 우선순위를 결정하기 위한 수학적모델로서 벡터스페이스모델(Vector Space Model)들과 병행하여 사용하고 있는 경우가 있다. 특히, 일반적인 검색엔진에서는 사용하는 빈도가 높았다.

46) Salton G. and McGill, M. J. 1983 Introduction to modern information retrieval. McGraw-Hill

벡터스페이스모델은 문헌과 용어에 대한 유사도와 우선순위를 결정하기 위한 수학적모델로서, 벡터의 다차원을 이용하여 중요도를 분석하기 위하여 사용하며 SMART(System for the Mechanical Analysis and Retrieval of Text, Cornell University, 1960)에 최초로 적용되어졌다.

이 모델은 특징은 용어에 대한 가중치를 주는 방식으로 기존의 불리언 모델보다 우수한 검색결과를 나타내었다. 이 모델의 원리는 문헌의 유사도는 용어에 대한 거리를 벡터에 의해 계산함으로써 구할 수 있으며, 용어에 대한 중요도는 TF-IDF모델을 통해 우선순위를 정할 수 있으며, 도출되는 결과는 문서중에서 가장 중요한 문서에 대한 우선순위를 가중치에 의해 구할 수 있으며, 그 문서내의 term의 중요도를 나타낼 수 있다.

$$\begin{aligned} k_i &: \text{용어} \\ d_i &: \text{문서} \\ w_{ij} &: \text{term Weight} \\ \vec{q} &= (w_{1q}, w_{2q}, \dots, w_{tq}), t = \text{term 빈도수} \\ \vec{d}_j &= (w_{1j}, w_{2j}, \dots, w_{tj}) \end{aligned}$$

$$\begin{aligned} \cos \theta &= \frac{\vec{d}_j \cdot \vec{q}}{|\vec{d}_j| |\vec{q}|} = \frac{(w_{1j} \times w_{1q} + \dots + w_{tj} \times w_{tq})}{\sqrt{w_{1q}^2 + \dots + w_{tq}^2} \sqrt{w_{1j}^2 + \dots + w_{tj}^2}} \\ C(d, q) &= \frac{(\sum_{i=1}^t w_{ij} \cdot w_{iq})}{\sqrt{\sum_{i=1}^t w_{iq}^2} \cdot \sqrt{\sum_{i=1}^t w_{ij}^2}} \end{aligned}$$

벡터스페이스모델의 단점은 term의 개수가 많아질수록 해당 문서는 유사도는 낮아지는 현상, 언어의 특성에 따라 찾고자하는 키워드가 정확히 매칭이 되지 않는 경우, 너무 광범위한 주제를 다루는 문서가 많으면 term의

양도 많아지기 때문에 비슷한 주제를 다루는 문서라도 전혀 다른 term을 도출하는 현상 등의 단점들이 나타나고 있지만, 최근 데이터베이스와 프로그램의 발달로 점점 더 오차가 줄어들고 있는 실정이다.

TF-IDF와 벡터스페이스모델을 활용한 사례를 다음과 같으며, 일반적으로 두 모델을 사용하여 중요도를 분석하는 방법이다.

TERM VECTOR MODEL BASED ON $W_i = t_{fi} \times IDF_i$											
<ul style="list-style-type: none"> <li>• Q : gold silver treckuery</li> <li>• D1 : 'Shipment of gold damaged in a fire'</li> <li>• D2 : 'Delivery of silver arrived in a silver truck'</li> <li>• D3 : 'Shipment of gold arrived in a truck'</li> <li>• <math>D=3</math>, <math>IDF = \text{LIG}(D/df_i)</math></li> </ul>											

Terms	Q	D1	D2	D3	df <sub>i</sub>	D/df <sub>i</sub>	IDF <sub>i</sub>	Q	D1	D2	D3
a	0	1	1	1	3	3/3=1	0	0	0	0	0
arrived	0	0	1	1	2	3/2=1.5	0.1767	0	0	0.1767	0.1767
damaged	0	1	0	0	1	3/1=3	0.4771	0	0.4771	0	0
delivery	0	0	1	0	1	3/1=3	0.4771	0	0	0.4771	0
fire	0	1	0	0	1	3/1=3	0.4771	0	0.4771	0	0
gold	1	1	0	1	2	3/2=1.5	0.1767	0.1767	0.1767		0.1767
in	0	1	1	1	3	3/3=1	0	0	0	0	0
of	0	1	1	1	3	3/3=1	0	0	0	0	0
silver	1	0	2	0	3	3/1=3	0.4771	0.4771	0	0.9542	0
shipment	0	1	0	1	2	3/2=1.5	0.1767	0	0.1767	0	0.1767
truck	1	0	1	1	2	3/2=1.5	0.1767	0.1767	0	0.1767	0.1767

$$|D_1| = \sqrt{0.4771^2 + 0.4771^2 + 0.1761^2 + 0.1761^2} = \sqrt{0.5173} = 0.7192$$

$$|D_2| = \sqrt{0.1761^2 + 0.4771^2 + 0.9542^2 + 0.1761^2} = \sqrt{1.200} = 1.0955$$

$$|D_3| = \sqrt{0.1761^2 + 0.1761^2 + 0.1761^2 + 0.1761^2} = \sqrt{0.1240} = 0.3522$$

$$\therefore |D_i| = \sqrt{\sum_i w_{i,j}^2}$$

$$|Q| = \sqrt{0.1761^2 + 0.4771^2 + 0.1761^2} = \sqrt{0.2896} = 0.5382$$

$$\therefore |Q| = \sqrt{\sum_i w_{Q,j}^2}$$

$$Q \bullet D_1 = 0.1761 \times 0.1761 = 0.0310$$

$$Q \bullet D_2 = 0.4771 \times 0.9542 + 0.1761 \times 0.1761 = 0.4862$$

$$Q \bullet D_3 = 0.1761 \times 0.1761 + 0.1761 \times 0.1761 = 0.0620$$

$$\therefore Q \bullet D_1 = \sum_i w_{Q,j} w_{i,j}$$

$$\cos \theta_{D_1} = \frac{Q \bullet D_1}{|Q| \bullet |D_1|} = \frac{0.0310}{0.5382 \times 0.7192} = 0.0801$$

$$\cos \theta_{D_2} = \frac{Q \bullet D_2}{|Q| \bullet |D_2|} = \frac{0.4862}{0.5382 \times 1.0955} = 0.8246$$

$$\cos \theta_{D_3} = \frac{Q \bullet D_3}{|Q| \bullet |D_3|} = \frac{0.0620}{0.5382 \times 0.3522} = 0.3271$$

$$\therefore \cos \theta_{D_3} = \sin (Q, D_I)$$

$$\therefore \sin (Q, D_I) = \frac{\sum_i w_{Q,j} w_{i,j}}{\sqrt{\sum_j w_{Q,j}^2} \sqrt{\sum_i w_{i,j}^2}}$$

중요도 랭킹

$$\text{Rank1:Doc2} = 0.8246$$

$$\text{Rank2:Doc3} = 0.3271$$

$$\text{Rank3:Doc1} = 0.0801$$

## 제7절 텍스트마이닝을 이용한 트렌드 분석방법

위와 같이 TF-IDF<sup>47)</sup>(Term Frequency - Inverse Document Frequency)는 트렌드분석이나 현황을 분석하는데 어려움이 존재한다. 이러한 현재의 트렌드를 분석하기 위하여 기존의 중요도를 보완하여 새로운 알고리즘이 필요하게 되었다.

새롭게 제시되는 TF-DI(Term Frequency - Date Index)는 미래의 트렌드를 분석하기 위한 텍스트마이닝 기법의 하나로써, 특정 키워드가 연도별로 얼마나 중요한지를 나타내는 가중치로 일종의 통계적 방법을 제시하고 있다. 이러한 방법은 TF-IDF의 변형하여 단점들을 보강하고 특정목적(트렌드 분석)을 위하여 개발되었다.

TF-DI의 가장 중요한 원리는 문서내의 중요한 키워드들을 도출하여 인터넷의 정보량에 따라 단어의 빈도수를 분석하는 방식을 사용한다. 또한, 문서군내의 특정단어의 문서간 빈도수를 사용하는 것이 아니라 연도별 가중치를 사용함으로써 트렌드 분석이 가능하도록 설계하였다. 즉 TF-IDF의 빈도수가 중요하다는 원리를 이용하지만, 트렌드를 분석하기 위하여 문서의 중요도는 인터넷을 활용한 연도별 가중치 분석으로 중요도를 분석하고 있다. 이러한 분석은 현재의 인터넷의 정보를 가중치로 이용하는 방법으로 현재의 데이터나 자료가 잘 반영되어 있다는 것이 장점이 있다.

TF-IDF와 TF-DI의 가장 큰 차이점은 첫째, TF-IDF 빈도수분석은 문서군들에 속해진 문서들의 빈도수를 이용하는 것이고, TF-DI은 중요하다고 생각되는 보고서, 논문 등의 중요문서들의 키워드들을 인터넷의 최근 검색 결과값을 들을 빈도수로 이용하는 것이다. 이 빈도수는 특정한 문서군에

47) Salton G. and McGill, M. J. 1983 Introduction to modern information retrieval. McGraw-Hill

국한되지 않으며, 최근의 트렌드를 반영하는 지표가 된다. 둘째, TF-IDF는 문서간 특정단어의 속한 문서들의 수에 따라 중요도를 판별하였으나, TF-DI에서는 시간의 개념을 변수로 사용하여 중요도를 산정하였다. 이는 시간개념을 도입함으로써, 최근까지의 트렌드를 분석할 수 있는 유용한 방법으로 제시하고 있다. 셋째, TF-IDF의 중요도는 어떤 문서군을 선택하는냐에 따라 편차가 심하게 발생하나, TF-DI는 인터넷을 이용한 결과값들을 사용하기 때문에 오차도 적을뿐더러 다양한 조건하에 값들을 중요도를 추출하고 트렌드를 분석할 수 있다

<표 3> TF-IDF와 TF-DI 방식 비교

구분	TF-IDF	TF-DI	비고
단어의 도출	-일정 문서군에서 문서내의 단어의 빈도수를 이용하여 TF를 구함(특정한 범위의 문서군을 설정하여야 함)	-가장 중요한 문서나 보고서 등을 어떤 형태를 가리지 않고 단어의 빈도수를 이용하여 우선순위들 도출함 -도출된 단어의 실제 빈도수를 인터넷의 검색 결과값을 이용하여 우선순위를 도출하였음..	-TF-DI 모델은 어떤 문서군을 형성할 필요 없이 중요하다고 생각되는 문서나 보고서 등을 다양하게 이용하여 단어들 도출
문서간 중요도 분석	-DI의 가중치는 문서간의 단어의 빈도수를 구함.	-DI는 트렌드 분석이 필요한 기간동안의 가중치를 이용함. -예를 들면, 최근 5년간의 트렌드를 분석하기 위한 DI는 최근 연도도 갈수록 가중치가 높음.	-TF-DI 모델은 시간의 개념이 포함되어 있음.
중요도 지수	-중요도 지수는 TF와 IDF를 곱해서 특정 단어의 중요도를 표시함.	-중요도 지수는 인터넷을 활용한 TF와 시간축을 중심으로한 DI를 곱해서 표시함	-시간별 가중치가 중요한 변수가 되므로, 시간별 트렌드 분석이 가능함.

이러한 접근방식의 차이에 의해 TF-IDF와 TF-DI는 중요도를 판별하는 유사한 스타일로 볼수 있으나, 시간의 개념이 포함되면서 완전히 다른 중요도를 도출하는 방법으로 인식된다. 다음 <표5>는 TF-IDF와 TF-DI의 장단점과 비교분석을 통해 비교우위를 분석하였다.

<표 4> TF-IDF와 TF-DF 방식의 장단점 분석

분석 내용	TF-IDF	TF-DF	비교분석
특성	특정 문서내의 단어를 추출하여 문서와의 관계에 따라, 단어의 우선순위를 도출이 가능함.	다양한 문서에서 단어를 추출하고 인터넷을 활용하기 때문에, 미래예측을 위한 트렌드 분석과 우선순위 도출이 가능함	TF-IDF는 검색엔진의 알고리즘으로 사용가능하나, TF-DF는 검색엔진의 결과를 활용함
사용용도	특정키워드 도출, 특정문서의 우선순위 도출,	특정키워드 도출, 미래 트렌드 분석을 위한 키워드 우선순위 도출	트렌드분석의 용이성 TF-IDF < TF-DF
자료의 수집 및 접근 용이성	문서내의 단어 빈도수는 문서의 종류와 문서군이 중요하며, 적절한 문서와 문서군이 아닌 경우 전혀 다른 결과를 초래하므로 적절한 너무 많은 시간과 경비를 소요함	문서를 기준으로 하지 않기 때문에 키워드를 추출하기 위해서 문서의 종류나 문서군에 제약을 받지 않고, 특정한 상황이나 조건을 가정하여 다양한 종류의 문서에서 필요한 키워드를 추출할 수 있음	자료수집의 용이성 TF-IDF < TF-DF
문서의 종류에 대한 효용성	문서내의 단어 빈도수는 논문, 특허 등 문서의 종류에 따라 많은 영향을 받음	인터넷을 통한 단어빈도수를 도출하기 때문에 문서의 종류나 영향을 많이 받지 않으므로, 보고서, 논문, 컨퍼런스 자료등 특정목적에 맞는 다양한 단어를 선택할 수 있음	문서의 제한성 TF-IDF > TF-DF
최신 트렌드 분석 가능 여부	미래예측시 최신 연구경향을 반영하기 위해 논문을 대상으로 하나, 최신 논문의 범위와 논문의 수를 판단하거나 분류하기가 어려움이 따르며, 또한 해당 전문가의 도움이 반드시 필요함	인터넷을 활용하기 때문에 최신 트렌드나 추세를 반영하기 쉬우며, 데이터 소스에서도 논문, 특허, 블로그 등 다양한 소스를 통해 똑같은 조건내에서 추세를 파악할 수 있음.	최신 트렌드 반영 TF-IDF < TF-DF



분석 내용	TF-IDF	TF-DF	비교분석
연도별 분석 용이성	TF-IDF 알고리즘상 연도별 분석이 거의 불가능하면, 만약 분석하기 위해서는 해당문서의 시간별로 분류해야하기 때문에 거의 불가능함	TF-DF는 연도별 빈도수가 발생하고 연도별 가중치를 염두하고 설계하였기 때문에 트렌드 분석 용이함	시계열분석 용이성 TF-IDF < TF-DF
분석의 오차 범위	문서의 내의 단어의 갯수가 많아질수록 해당 문서는 더 낮은 similarity 값을 가지게 되므로, 문서의 크기가 클수록 저평가가 될 확률이 높음	특정 키워드의 단어수가 아무리 많아도 별로 영향을 받지 않고, 단어가 많을수록 단어별로 분석하기가 용이함.	분석의 오차 TF-IDF > TF-DF
대상문서의 범위의 한계	위낙에 광범위한 주제를 다루는 문서를 대상으로 하는 경우에는 문서에 포함된 단어의 양도 많아지고, 비슷한 주제를 다루는 문서라고 하더라도 전혀 다른 단어구성을 나타낼 확률이 높음	각 단어별로 연도 및 개별 분석이 가능하고, 특정 주제를 묶어서 분석이 가능하기 때문에,	다양한 키워드 분석 TF-IDF < TF-DF
실제 활용도와 및 시스템 개발의 용이성	문서가 많아질수록 계산하는 과정이 복잡하고 느려지고, 시스템개발이 많은 데이터베이스와 인덱스를 사용하므로 시스템 개발시간과 능	구조가 간단하고 심플하여 계산과정이 속도가 빠르며, 구글의 검색 Open API를 활용하기 때문에 유연한 시스템개발이 가능함.	활용도 및 개발용이성 TF-IDF < TF-DF

이러한 TF-DI(Term Frequency - Date Index)의 세부적인 알고리즘을 살펴보면 2가지의 명제를 중심으로 개발되었다.

#### <명제>

1. 문서나 프로세스에 의해 도출된 키워드는 빈도수가 높은 것이 중요하다.
2. 연도별 키워드 빈도가 높다는 것은 중요한 키워드이다.

첫 번째 명제는 특정 키워드의 빈도수가 높다는 것은 특정 키워드가 중

요하다는 것을 의미하므로, 인터넷의 노출 빈도가 높은 키워드를 빈도수로 선정하였다. 즉, 인터넷 검색엔진에 의한 검색결과 값을 빈도수로 선정하였다.

두 번째 명제는 특정키워드의 빈도수가 최근으로 올수록 중요한 요소라고 예상하여, DF(Date Frequency)에서는 키워드의 연도별 빈도수와 연도별 가중치에 부여하여 산정하였다(최근년도로 갈수록 가중치가 높음)

이러한 두가지 명제를 중심으로 TF-DI(Term Frequency - Date Index)에서는 키워드의 상대적 빈도수와 연도별 가중치의 곱을 표현하여 아래와 같은 수식을 도출하였다.

$$(TF-DI)_i = \sum_{j=1}^n TF_j \times DI_j$$

$$TF_j = \frac{i - frequency}{(total frequency)_j}$$

$$DI_j = \frac{j}{n}$$

$i$  = 키워드 구분

$j$  = 시작되는 연도의 첫 횟수( $j=1$ )

$n$  = 분석기간의 연도별 횟수( $j=1, 2, 3 \dots$ )

예를 들면, 특정문서에 중요한 키워드 A, B, C가 있다면 가정하면 아래 그림과 같이 검색엔진을 이용하여 A, B, C의 키워드들의 검색결과 값을 도출하였다.



<그림 11> 검색엔진을 이용한 검색결과값

검색범위는 2007~2010년 4년동안의 각각의 키워드들 검색하였으며 빈도수를 나타낸 것이다. 일반적으로 키워드 총 빈도수에 의해서는  $A > C > B$ 의 순으로 가장 많은 빈도수를 나타냈으며, 이를 중요도가 큰 순서로 표현한 것이다.

빈도수	2007	2008	2009	2010	총 빈도합
A키워드	340	409	500	300	1,549
B키워드	210	300	420	510	1,440
C키워드	110	600	400	400	1,510
빈도합(년도)	660	1,309	1,320	1,210	

TF-DI(Term Frequency - Date Index)에서 TF는 구글의 검색엔진에서 검색되어진 결과값을 빈도수로 가정하고 모든 키워드의 상대적 빈도수를 표현하여 각 연도별로 비교가 가능하도록 도출하였다.

또한, 트렌드분석시 과거연도에서 최근 연도가 올수록 더 중요하다는 일

반적인 명제에 의해 DF는 분석하려는 해당 총 연도의 역수에 차년도의 연도를 곱하여 최근 연도로 올수록 가중치가 높도록 보정한 식이다(예를 들면, 2005년부터 2007년까지 분석할 경우, 2005년  $DF=1/3$ , 2006년  $DF=2/3$ , 2007년  $DF=3/3$ ).

따라서 TF-DI에서 TF와 DI를 곱하여 다음과 같은 중요도를 분석하였다. 단순한 빈도수의 분석에서는  $A>C>B$ 의 순으로 나왔으나, 최종적인 TF-DI 중요도에서는  $B>C>A$ 순으로 나타났다.

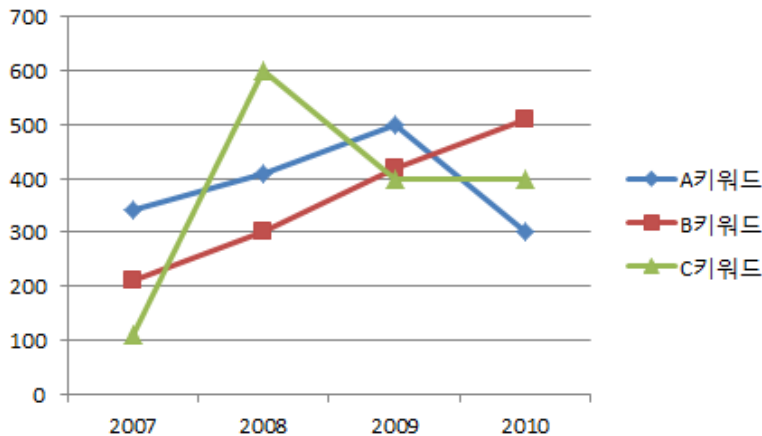
가중치	2007	2008	2009	2010	TF-DI
A키워드	0.1287879	0.1562261	0.2840909	0.2479339	0.8170388
B키워드	0.0795455	0.1145913	0.2386364	0.4214876	0.8542607
C키워드	0.0416667	0.2291826	0.2272727	0.3305785	0.8287005

이는 아래그림의 A, B, C의 절대적 빈도수를 비교한 곡선이다. 절대적 빈도수에서는 A키워드가 가장 높으나 그림에서는 최근년도로 올수록 하강 곡선을 나타내고 있다.

C키워드는 초창기 상향 성장곡선을 나타내다가 최근에 와서 성장곡선을 유지하고 있다.

B키워드는 전체 절대적인 빈도수에서는 가장 작은 빈도수를 나타냈지만, 최근까지 지속적인 성장곡성을 나타내고 있다.

이와 같이, TF-DI(Term Frequency - Date Index)의 결과에서도  $B>C>A$ 순으로 나타났으며, 그림과 같이 분석할 수 있는 중요도를 수식으로 잘 표현하고 있다.



<그림 12> 년도별 키워드의 절대적 빈도수

#### 제8절 새로운 복합적 미래예측 방법

본 연구에서는 미래를 예측하는 정량적인 방법으로 본 연구에서는 텍스트 마이닝과 복잡계 네트워크를 이용한 방법을 이용하였다. 그러나, 이와 같은 기존의 방법들은 각각 다른 방향으로 존재하는 방식으로 미래를 예측하는 정량적인 방법으로 활용하고 적응하기에는 어려움이 따랐다. 따라서 본 연구에서는 기존의 방식을 개선하여 중요한 키워드의 우선순위를 도출하였다. 특히, 텍스트마이닝 방법에서는 방대한 정보에서 현재의 트렌드를 반영한 중요한 키워드들을 도출하고, 복잡계 네트워크를 통해서 도출된 키워드들간의 상호연관관계를 통해 우선순위를 도출하였다.

기존의 텍스트마이닝 기법중에 하나인 TF-IDF(Term Frequency - Inverse Document Frequency)는 문서내의 단어 빈도수를 이용하여 키워드를 도출하였으나, 문서의 종류와 문서군에 따라 상이한 결과가 나타나고 있으며 너무 많은 시간과 경비를 소요한다. 또한, 최근의 트렌드를 반영하는 키워드를 도출하는데 많은 어려움이 따르고 있다.

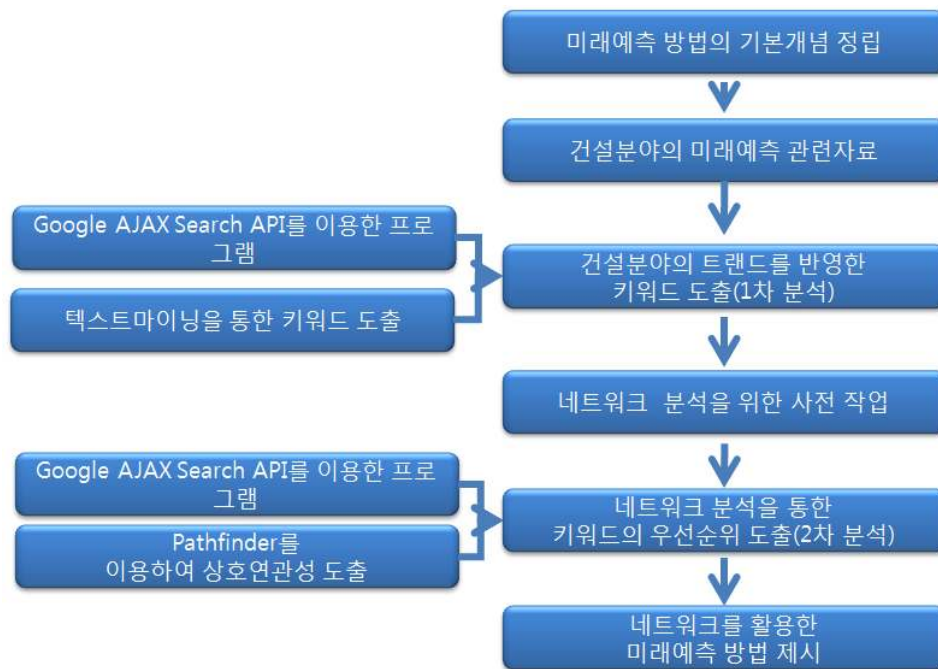
본 연구에서는 기존의 TF-IDF(Term Frequency - Inverse Document Frequency)가 방대한 정보에서 중요한 키워드들을 도출하는 기본적인 개념을 중심으로 새로운 알고리즘을 도입하였다. 제시된 TF-DI(Term Frequency - Date Index)는 문서내에서 반복적인 노출된 키워드들을 빈도수에 따라 나열하여 새로운 키워드 그룹을 형성하고, 이 키워드들의 인터넷에 노출된 정보 검색결과량을 바탕으로 키워드의 빈도수를 분석하는 방식을 사용하였다. 즉, 문서군내의 특정단어의 문서간 빈도수를 사용하는 것이 아니라 현재 인터넷의 정보를 이용한 검색결과값을 빈도수로 이용하고 연도별 가중치를 사용함으로써 트렌드 분석이 가능하도록 설계하였다. 이러한 텍스트마이닝 기법을 활용하여 방대한 정보로부터 현재 트렌드를 반영하는 중요한 키워드를 1차적으로 추출하였다.

인터넷의 정보를 반영하기 위해서 구글에서 제공하는 Open API(Application Programming Interface)를 사용하여 검색결과를 얻었으며, 검색기간, 다중 검색어 검색 등의 다양한 기능들을 활용하였다. 이렇게 도출되어진 키워드들은 복잡계 네트워크 방식을 활용해 키워드간의 상호연관관계를 분석하여 우선순위를 도출하였다. 이러한 우선순위를 도출하기 위하여 복잡계의 네트워크를 이용한 중심성 분석(Centrality Analysis)과 네트워크의 간략화 및 최적화 기법을 이용한 2가지 방식을 결합하여 분석하였다.

복잡계 네트워크의 중심성 분석방법은 각각의 행위자를 표현하는 노드와 상호 연관관계를 의미하는 링크의 상호관계를 규명하는 중요한 방식이다. 특히, 어떤 노드가 허브 역할을 하는지, 어떤 노드가 중계자의 역할을 하는지 등의 중심에 대한 해석이 가장 중요하며, 전체 네트워크에서 각각의 노드들에 대한 역할, 위치, 특성, 영향력을 파악할 수 있다. 또한, 네트워크 내에서 노드들의 상대적 위치나 절대적 위치를 파악하는데 유용하게 활용할 수 있다.

이러한 네트워크의 중심성을 분석하기 위해서는 노드간의 어떤 상호연관 관계가 필요하다. 그러나, 현재까지 기술간 상호연관관계는 결정하기 위하여 일반적으로 전문가들의 정성적인 평가에 의해 결정되었다. 따라서, 본 연구에서는 정량적인 상호연관관계를 분석하기 위하여, 현재 기술들이 인터넷에 얼마만큼 노출된 정도(인터넷에 노출된 검색 결과값)에 따라 네트워크간의 상호관계를 결정하게 된다.

또한, 이렇게 작성된 네트워크는 노드 수  $n$ 에 대해  $n(n-1)$ 만큼 링크수가 존재하기 때문에 네트워크 구조 분석 시 링크가 너무 많아 중심성 분석에 많은 어려움이 있으며 시각적으로 표현하는데 한계가 있다. 따라서, 본 연구에서는 전체 네트워크에서 중요한 역할을 담당하는 링크만을 남겨두고 나머지를 제거하는 방법으로 간략화 및 최적화 방법을 활용하여 네트워크를 구성하게 된다. TF-DI(Term Frequency - Date Index)와 구글 API를 활용한 검색결과 프로그램을 이용한 텍스트마이닝 기법과 구글 API를 활용한 검색결과 프로그램을 이용한 상호연관관계와 네트워크의 간략화 및 최적화, 복잡계를 네트워크를 활용한 분석 등의 다양한 방법들을 결합하여, 방대한 자료에서 키워드의 도출과 우선순위를 분석하는 정량화 방법들을 본 연구에서 개발하여 적용하였다.



<그림 13> 새로운 복합적 미래예측 방법의 과정



## 제4장 건설분야에의 시범적 적용

### 제1절 제3회 과학기술예측조사

텍스트마이닝은 광의적으로 정의하면 구조화되지 않은 대규모 집합체로부터 새로운 지식이나 정보를 추출하는 일련의 과정을 말하며, 일반적으로는 다양하고 정형화 되어 있지 않는 문서와 데이터로부터 유용한 정보를 추출 및 변환하는 일련의 과정과 방법을 지칭한다. 텍스트마이닝 과정은 [비정형 정보수집 → 정보처리 → 정보추출 → 정보분석] 등의 일반적인 절차를 따르고 있다.

정보수집 과정에서는 과학기술예측조사와 관련 분야의 비정형 데이터, 기술자료, 논문 등의 자료를 수집하는 단계에서는 과학기술예측과 관련된 주요 키워드, 개요서 등의 자료 수집하였다.

정보처리 과정에서는 도출된 대상 자료와 기술리스트를 중심으로 각 기술에 해당되는 정보를 추출할 수 있는 데이터 소스를 가공하는 단계로 수집된 자료(문서)를 기본으로 관련 키워드의 추출과 추출된 키워드의 수정 및 검토를 통한 영문화하였다.

정보추출 과정에서는 수학적 모델(알고리즘)을 활용한 키워드 추출단계로 미래 키워드를 예측하는 모델에 적용하고 모델적용에 따른 주요 키워드 도출하였다.

#### ■ 정보수집 단계

미래예측 키워드 도출을 위하여 ‘제3회 과학기술예측조사 수정·보완’의 기술중 건설·교통분야의 11개의 중점과학기술에 따른 42개의 중점과학기술

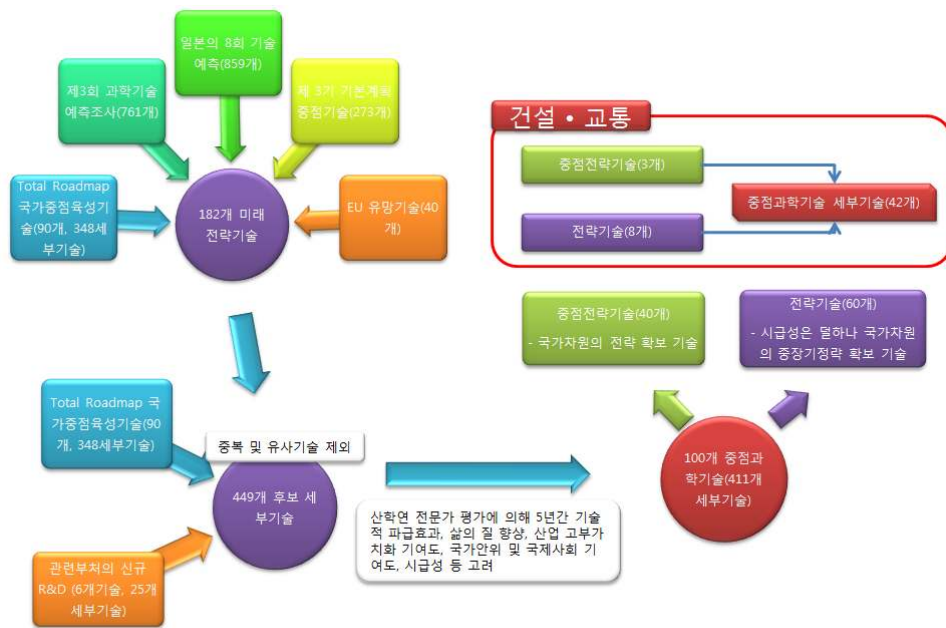
세부기술 선정하였다.

1차적으로, 제3회 과학기술예측조사(761개), Total Roadmap 국가중점육성기술(90개, 348개 세부기술), 일본의 8회 기술예측(859개), 제3기 기본계획 중점기술(273개), EU 유망기술(40개)<sup>48)</sup>에서 도출한 182개의 미래전략기술과 Total Roadmap 국가중점육성기술(90개, 348개 세부기술)과 관련부처의 신규 R&D 사업과제(6개, 25개 세부기술)을 바탕으로 449개의 후보 세보기술 도출하였다.

2차적으로, 산학연 전문가 평가에 의해 5년간 기술적 파급효과, 삶의 질 향상, 산업 고부가치화 기여도, 국가안위 및 국제사회 기여도, 시급성 등 고려하여 449개의 후보기술 중 100개의 중점과학기술과 411개의 세부기술로 선정하였다. 100개의 중점과학기술에서는 국가차원의 전략을 확보하기 위한 중점전략기술(40개)과 중점전략기술보다는 시급성은 덜하나 국가차원에서 중장기전략을 위하여 확보해야 되는 전략기술(60개)로 구분하였다. 이 과정에서 건설·교통과 관련된 중점전략기술 3개와 전략기술 8개를 대상으로 중점과학기술 세부기술 42개를 선정하였다.

---

48) Emerging Science and Technology priorities in public research policies in the EU, the US and Japan, European Commission, 2006



&lt;그림 14&gt; 건설·교통분야 핵심기술리스트 도출과정

&lt;표 5&gt; 중점과학기술 세부기술 목록 및 개요

기술 분야	중점 과학기술	no.	중점 과학기술 세부기술
건설·교통	초고층빌딩 건축기술	1	초고층 건축물 구조시스템 및 고성능 재료 기술
		2	초고속 시공 및 공사관리 기술
		3	Pilot Project 계획 및 설계기술
		4	고성능 환경 및 M&E 요소 기술
	건설기반기술	5	건설 생산성·안전성 향상 기술
		6	에너지/자원 절약형 건설기술
		7	첨단 건설재료/소재 개발 기술
		8	건설기자재 품질인증·검사·시험 기술
		9	신 건설생산시스템 및 건설표준화기술
		10	지능형 시설물 유지관리 기술
		11	대심도 지하공간 구축·활용기술

기술 분야	중점 과학기술	no.	중점 과학기술 세부기술
	초장대교량 건설기술	12	초장대 교량 설계기준 및 설계기법
		13	내진/내풍에 의한 거동예측 및 진동저감기술
		14	첨단 시공기술 및 특수 장비 개발기술
		15	초고성능 재료 개발 기술
	미래 첨단 도시건설 기술	16	U-Eco City (친환경 도시건설 기술)
		17	복합공간 개발 기술
		18	도시 재생 시스템 기술
	지능형 국토지리정보구축기술	19	국토 모니터링 기술
		20	공간정보 기반 인프라 기술
		21	U-GIS 기반 건설정보화 기술
	미래 첨단 주거·교육환경 기술	22	인텔리전트 교육시설 (강의실, 도서관 등) 건설기술
		23	친실버 및 친환경 주거환경 구현 기술
	미래 첨단 교통시스템기술	24	ITS 기술 (스마트 하이웨이기술, 교통체계효율화 기술, 교통사고 최소화 위한 도로안전성향상기술)
		25	U-Transportation 기술 (도로교통안전기술, 교통사고 예방을 위한 교통운영기술, 재난, 재해 대응형 교통시스템기술 등)
		26	미래 교통안전향상 기술 (교통정보기술, 교통시설기술, 교통관제기술, 차량안전도평가기술, 차량안전기술, 자동단속기술, 교통안전시설기술 등)
		27	교통연계 및 환승시스템 기술 (교통흐름관리기술, 전자지불기술 등)
	해양·항공운항 효율화 및 안전향상기술	28	중소형항공기 인증기술
		29	항공 교통·운항 효율화 기술
		30	항공안전기술
		31	해양안전기술
	차세대 고속열차 기술	32	해양교통관리기술
		33	400km/h 분산형 고속철도기술
	첨단도시형 철도시스템 기술	34	초고속자기부상열차기술
		35	첨단경량전철 및 신교통시스템기술
		36	도시형 자기부상 시스템 기술

기술 분야	중점 과학기술	no.	중점 과학기술 세부기술
	첨단물류기술	37	통합 지능형 컨테이너 시스템 기술
		38	u-통관 시스템기술 (유비쿼터스 통합 물류정보화 기술)
		39	지능형 교통물류 정보화 기술
		40	지능형 물류센터 및 고효율장비 기술
		41	미래형 수송시스템 기술
		42	해운물류기술

### ■ 정보처리 단계

11개의 중점과학기술의 개요서를 중심으로 42개의 중점과학기술 세부기술에 대한 미래관련 키워드들을 도출(별첨2 참조)하였다.

정보처리 과정에서는 도출된 대상 자료와 기술리스트를 중심으로 각 기술에 해당되는 정보를 추출할 수 있는 데이터 소스를 가공하는 단계로 수집된 자료(문서)를 기본으로 관련 키워드의 추출과 추출된 키워드의 수정 및 검토를 통한 영문화하였다.

&lt;표 6&gt; 기술 리스트 및 영문 키워드

기술 분야	중점 과학 기술	no.	중점 과학기술 세부기술	영문 키워드
건설·교통	초고층빌딩 건축기술	1	초고층 빌딩시스템 및 고성능 재료 기술	Skyscraper Skyscraper structure Skyscraper material
		2	초고속 시공 및 공사관리 기술	Skyscraper construction management
		3	Pilot Project 계획 및 설계기술	Skyscraper planning Skyscraper design
		4	고성능 환경 및 M&E 요소 기술	Skyscraper environment
	건설기반기술	5	건설 생산성·안전성 향상 기술	intelligent facility maintenance facility safety management
		6	에너지/자원 절약형 건설기술	green building
		7	첨단 건설재료/소재 개발 기술	advanced construction material
		8	건설기자재 품질인증·검사·시험 기술	construction quality certification
		9	신 건설생산시스템 및 건설표준화기술	Construction Standard construction production system
		10	지능형 시설물 유지관리 기술	—
		11	대심도 지하공간 구축·활용기술	deep underground
	초장대교량 건설기술	12	초장대 교량 설계기준 및 설계 기법	long-span bridge long-span bridge design
		13	내진/내풍에 의한 거동예측 및 진동저감기술	long-span bridge structure
		14	첨단 시공기술 및 특수 장비 개발기술	long-span bridge equipment long-span bridge management
		15	초고성능 재료 개발 기술	long-span bridge advanced material
	미래 첨단 도시건설 기술	16	U-Eco City (친환경 도시건설 기술)	green city
		17	복합공간 개발 기술	complex spatial
		18	도시 재생 시스템 기술	urban regeneration
	지능형 국토지리정보 구축기술	19	국토 모니터링 기술	Land Monitoring
		20	공간정보 기반 인프라 기술	spatial information
		21	U-GIS 기반 건설정보화 기술	Ubiquitous GIS

기술 분야	중점 과학 기술	no.	중점 과학기술 세부기술	영문 키워드
	미래 첨단 주거·교육환경 기술	22	인텔리전트 교육시설건설기술	intelligent education facility
		23	친실버 및 친환경 주거환경 구현 기술	silver housing environment-friendly housing
	미래 첨단 교통시스템 기술	24	ITS 기술	intelligent transportation system
		25	U-Transportation 기술	Ubiquitous Transportation
		26	미래 교통안전향상 기술	Transportation safety
		27	교통연계 및 환승시스템 기술	Integration Transportation
	해양·항공운항 효율화 및 안전향상기술	28	중소형항공기 인증기술	small aircraft certification
		29	항공 교통·운항 효율화 기술	Air Transport efficiency
		30	항공안전기술	Aviation Safety
		31	해양안전기술	Marine safety
		32	해양교통관리기술	marine transportation management
	차세대 고속열차 기술	33	400km/h 분산형 고속철도기술	high-speed rail
		34	초고속자기부상열차기술	magnetic levitation train
	첨단도시형 철도시스템 기술	35	첨단경량전철 및 신교통시스템 기술	light rail
		36	도시형 자기부상 시스템 기술	-
	첨단물류기술	37	통합 지능형 컨테이너 시스템 기술	intelligent logistics
		38	u-통관 시스템기술	ubiquitous logistics
		39	지능형 교통물류 정보화 기술	-
		40	지능형 물류센터 및 고효율장비 기술	logistics automation
		41	미래형 수송시스템 기술	-
		42	해운물류기술	shipping logistics

## ■ 정보추출 단계

정보처리 단계에서 도출된 영문키워드를 중심으로 TF-DF(Term Frequency - Date Frequency)를 활용하여 키워드의 가중치 도출하였다. 검색은 구글 검색엔진을 활용하여 검색 결과값을 가중치로 도출하였다. 검색기간은 2000.1.1. ~ 2009.12.31.(10년간)로 설정하여 TF-DF(Term Frequency -

Date Frequency)에 의한 주요 키워드 선정하였다. 특히, 카테고리가 크거나 대표성이 있는 기술은 제외하고 같은 기술의 여러 가지 세부기술들은 우선순위가 높은 기술로 선정하였다(예, 초고층 빌딩 설비 및 환경, 초고층 빌딩 설계, 초고층 빌딩 계획 등은 중요도 지수가 높은 것을 선정). 또한, 다른 기술에 비해 TF-DF 너무 낮은 것은 제외하였다(예, ubiquitous 물류, 공간정보 기반 인프라 기술, 친환경 주거, U-Transportation 기술, 중소형항공기 )

<표 7> 중요도 지수로(TI)로 도출된 기술 리스트 및 영문 키워드

	중점과학 세부기술	영문키워드	TF-DF	선정여부
1	초고층 빌딩	Skyscraper	11290.114	X
2	경량전철	light rail	5206.1173	
3	초고층 빌딩 설비 및 환경	Skyscraper environment	3108.8757	
4	초장대 교량 장비	long-span bridge equipment	2792.2539	
5	초고속자기부상열차기술	magnetic levitation train	2616.0213	
6	미래 교통안전향상 기술	Transportationsafety	2564.0572	
7	초고층 빌딩 설계	Skyscraper design	2404.26	X
8	항공안전기술	Aviation Safety	2252.0875	
9	시설안전	facility safety management	2087.2799	
10	400km/h 분산형 고속철도기술	high-speedrail	1901.2445	
11	지능형 시설 유지관리	intelligent facility maintenance	1607.9018	
12	국토 모니터링 기술	LandMonitoring	1538.7327	
13	초장대 교량 설계	long-span bridge design	1482.1231	X
14	해양교통관리기술	marine transportation management	1462.6673	
15	초장대교량재료	long-span bridge advanced material	1357.7789	X
16	복합공간개발기술	complex spatial	1197.1244	
17	항공 교통·운항 효율화 기술	Air Transport efficiency	1107.9809	
18	U-GIS 기반 건설정보화 기술	Ubiquitous GIS	1090.3	
19	초고층 빌딩 계획	Skyscraper planning	963.1944	X



	중점과학 세부기술	영문키워드	TF-DF	선정여부
20	대심도 지하공간	deepunderground	796.05731	
21	해양안전기술	Marine safety	795.93804	
22	인텔리전트 교육시설건설기술	intelligent education facility	692.78602	
23	ITS 기술	intelligent transportation system	673.7842	
24	초장대 교량	long-span bridge	514.95725	X
25	초장대 교량 구조	long-span bridge structure	505.10814	X
26	건설생산시스템	construction production system	454.95227	
27	도시 재생 시스템 기술	urbanregeneration	454.27505	
28	첨단재료/소재	advanced construction material	420.49649	
29	건설인증	construction quality certification	391.69061	
30	해운물류기술	shippinglogistics	336.51971	
31	건설표준화	Construction Standard	195.91045	
32	에너지/자원절약형건설기술	green building	186.51108	
33	초장대 교량 건설관리	long-spanbridgemanagement	109.88248	X
34	실버 주거환경	silver housing	104.33369	
35	친환경 도시건설 기술	green city	82.691987	
36	물류 자동화	logistics automation	64.618014	
37	지능형 물류	intelligent logistics	50.574406	
38	교통연계 및 환승시스템 기술	Integration Transportation	37.161948	X
39	ubiquitous 물류	ubiquitous logistics	29.538344	X
40	초고층 빌딩 시공 및 공사관리	Skyscraper construction management	25.117828	X
41	공간정보 기반 인프라 기술	spatial information	14.310017	X
42	친환경 주거	environment-friendly housing	13.049058	X
43	초고층 빌딩 재료	Skyscraper material	9.047507	X
44	초고층 빌딩구조	Skyscraper structure	6.665581	X
45	U-Transportation 기술	Ubiquitous Transportation	2.5495492	X
46	중소형항공기 인증기술	smallaircraftcertification	1.3581987	X

## ■ 정보분석 단계

정보분석 단계에서는 중점과학기술의 카테고리별로 그룹핑된 기존의 영문키워드와 새로 도출된 키워드를 분석하여 새로운 연관성이 있는 기술들의 그룹명과 그룹핑을 실시하였다. 중점과학기술 세부기술을 기술리스트로 변경하고 그룹명을 재수정하였다.

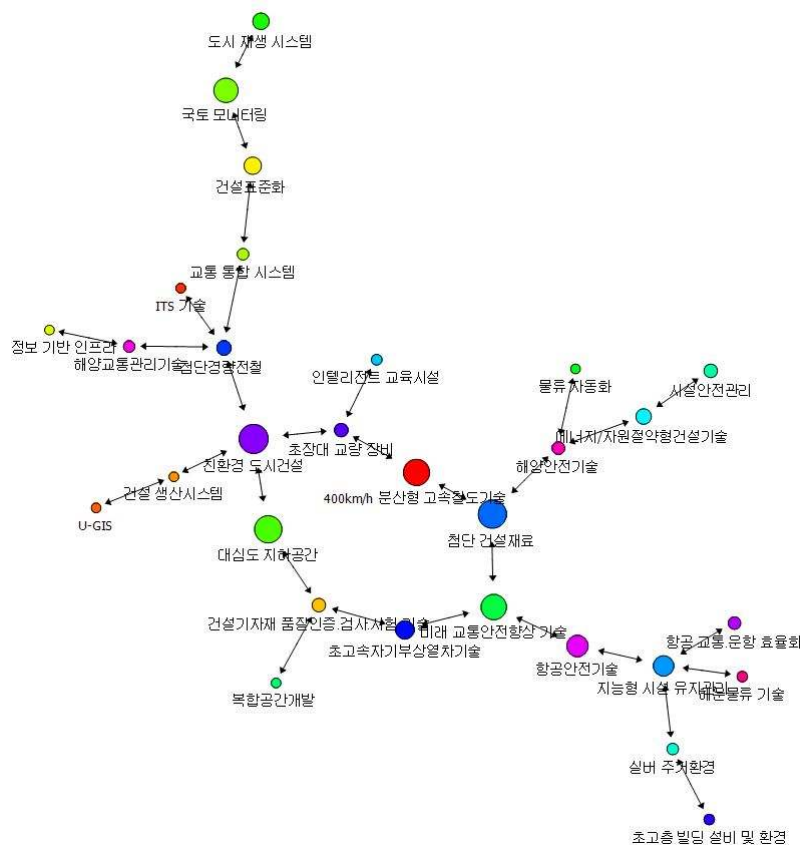
<표 8> 최종 도출된 기술 리스트 및 영문 키워드

중분류	no	기술리스트	영문키워드
건설 기반기술	1	지능형 시설 유지관리	intelligent facility maintenance
	2	시설안전관리	facility safety management
	3	에너지/자원절약형건설기술	green building
	4	첨단 건설재료	advanced construction material
	5	건설기자재 품질인증검사시험 기술	construction quality certification
	6	건설표준화	Construction Standard
	7	건설 생산시스템	construction production system
거대 구조물	8	초고층 빌딩 설비 및 환경	Skyscraper facility
	9	대심도 지하공간	deep underground
	10	초장대 교량 장비	long-span bridge equipment
	11	복합공간개발	complex spatial
미래 철도	12	초고속자기부상열차기술	magnetic levitation train
	13	400km/h 분산형 고속철도기술	high-speed rail
	14	첨단경량전철	light rail
효율적인 국토관리	15	U-GIS	Ubiquitous GIS
	16	국토 모니터링	Land Monitoring
	17	공간정보 기반 인프라	spatial information
항공 및 해양	18	항공 교통·운행 효율화	Air Transport efficiency
	19	항공안전기술	Aviation Safety
	20	해양안전기술	Marine safety
	21	해양교통관리기술	marine transportation management
미래 교통	22	ITS 기술	intelligent transportation system
	23	미래 교통안전향상 기술	Transportation safety
	24	교통 통합 시스템	Integration Transportation
미래 주거환경	25	친환경 도시건설	green city
	26	도시 재생 시스템	urban regeneration
	27	실버 주거환경	silver housing
	28	인텔리전트 교육시설	intelligent education facility
첨단 물류	29	해운물류 기술	shipping logistics
	30	물류 자동화	logistics automation

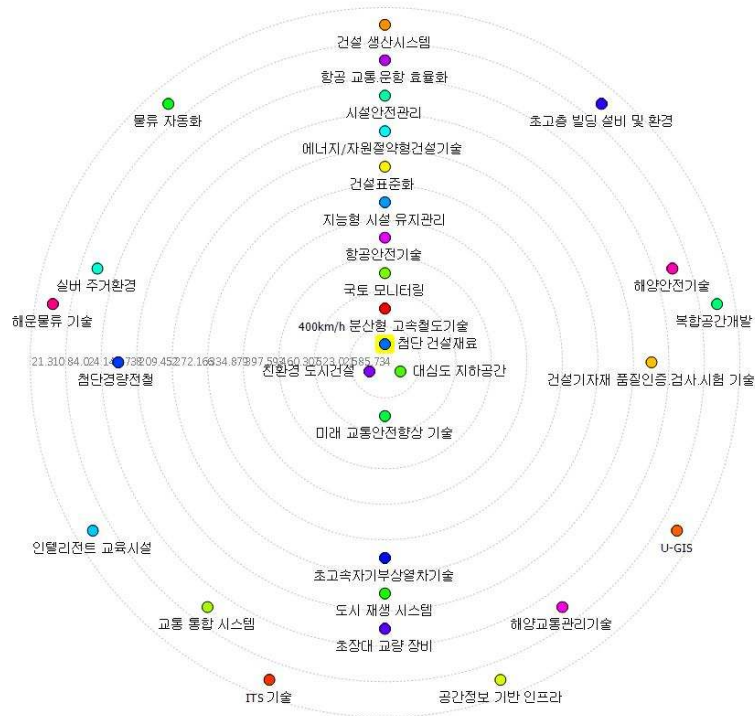
## ■ 네트워크 분석단계

건설분야의 네트워크 분석하기 위하여, 다음과 같은 연구과정을 수행하였다.

- 건설분야 30개 요소기술에 대한 노드로 선정
- Google AJAX Search API를 이용하여 요소기술간 상호 연관관계 도출(검색결과값을 이용한 가중치 선정)
- 패스파인더방식(Pathfinder : PFNet) 방식을 이용하여 비방향성의  $n(n-1)/2$  링크를 가중치의 중요도에 따라 간소화 및 최적화



<그림 15> 건설분야 네트워크 분석-I



<그림 16> 건설분야 네트워크 분석-II

## 제2절 과학기술 미래비전

2010.2월에 발표된 과학기술 미래비전의 내용중 건설분야와 관련된 내용을 추출하면 다음과 같다.

### - 4대분야의 미래모습 선정

- 자연과 함께하는 세상
- 풍요로운 세상
- 건강한 세상
- 편리한 세상

- 편리한 세상의 6대 트렌드 중 건설분야의 관련된 6개중 2개의 트렌드
  - 복합공간과 생태도시 개발 기술은 육상 공간의 과밀화와 생활 패턴 변화 등에 대응하기 위해서 점차 그 중요성이 높아질 것이다.
  - 새로운 물류·운송 수단의 등장으로 교통수단의 효율성이 증대되어 이동 시간이 단축되고 생활권이 확대될 것이다.

## ■ 정보처리 단계

시나리오에 의해 작성된 미래 트렌드에서 주요 키워드 도출하였다. 추출된 키워드의 수정 및 검토를 통한 영문화하였다.

<표 9> 시나리오에 의한 주요 기술목록

No	분류	기술명(한글)	기술명(영문)
1	건설	대규모 구조물	large structure
2		지하공간	deep underground
3		인공섬	Artificial island
4		해저터널	undersea tunnel
5		해양도시	ocean city
6		첨단 폐기물 시설	Waste Manifest System
7		초고층 빌딩	Skyscraper
8		도시방재	urban disaster prevention
9		시설물 안정성	facility safety
10		국토공간정보 활용	spatial information
11		무인 시공현장	automated construction
12		건축물 이동 기술	building motion
13		우주기지 건설	space station construction
14	교통	무인 자동 운전 시스템	unmanned vehicles
15		대심도 지하도로	deep road
16		대심도 지하철도	deep railroad

No	분류	기술명(한글)	기술명(영문)
17		자기부상열차	magnetic levitation train
18		수소자동차	hydrogen car
19		연료전지 교통기관	fuel cell transportation
20		첨단 ITS	intelligent transportation system
21		첨단 자동차 안전기술	Transportation safety
22		파일럿 정보시스템	Head Up Display(HUD)
23		실시간 차량정보 시스템	Vehicle Information Communication System(VICS)
24		개인선택형 대중교통	Personal Rapid Transit(PRT)
25		지능형 화물 운송 및 관리	intelligent logistics
26		유비쿼터스 통합물류 정보기술	ubiquitous logistics
27		진공터널 등의 고속 물류운송 튜브망	tube transportation
28		대량 물류 수송수단	large transportation
29		우주 수송기	space transportation

## ■ 정보추출 단계

정보처리 단계에서 도출된 영문키워드를 중심으로 TF-DF(Term Frequency - Date Frequency)를 활용하여 키워드의 가중치 도출하였다. 방법은 구글 검색엔진을 활용하여 검색 결과값을 가중치로 도출하였으며, 검색기간은 2000.1.1. ~ 2009.12.31. (10년간)을 대상으로 하였다.

&lt;표 10&gt; 주요기술의 TF-DF

No	분류	기술명(한글)	기술명(영문)	TF-DF
1	건설	대규모 구조물	large structure	0.1015096
2		지하공간	deep underground	0.5906837
3		인공섬	Artificial island	0.0588668
4		해저터널	undersea tunnel	0.0117184
5		해양도시	ocean city	1.0999907
6		첨단 폐기물 시설	Waste Manifest System	0.0013735
7		초고층 빌딩	Skyscraper	1.8556025
8		도시방재	urban disaster prevention	0.0026285
9		시설물 안정성	facility safety	0.0479931
10		국토공간정보 활용	spatial information	0.6536529
11	교통	무인 시공현장	automated construction	0.0271896
12		건축물 이동 기술	building motion	0.0101684
13		우주기지 건설	space station construction	0.0047533
14		무인 자동 운전 시스템	unmanned vehicles	0.0594938
15		대심도 지하도로	deep road	0.004304
16		대심도 지하철도	deep railroad	0.0005029
17		자기부상열차	magnetic levitation train	0.0109483
18		수소자동차	hydrogen car	0.0544719
19		연료전지 교통기관	fuel cell transportation	0.0013928
20		첨단 ITS	intelligent transportation system	0.0991385
21		첨단 자동차 안전기술	Transportation safety	0.439549
22		파일럿 정보시스템	Head Up Display(HUD)	0.169426
23		실시간 차량정보 시스템	Vehicle Information Communication System(VICS)	0.0010955
24		개인선택형 대중교통	Personal Rapid Transit(PRT)	0.026965
25		지능형 화물 운송 및 관리	intelligent logistics	0.007569
26		유비쿼터스 통합물류 정보기술	ubiquitous logistics	0.004052
27		진공터널 등의 고속 물류운송 튜브망	tube transportation	0.0050366
28		대량 물류 수송수단	large transportation	0.0371752
29		우주 수송기	space transportation	0.1127485

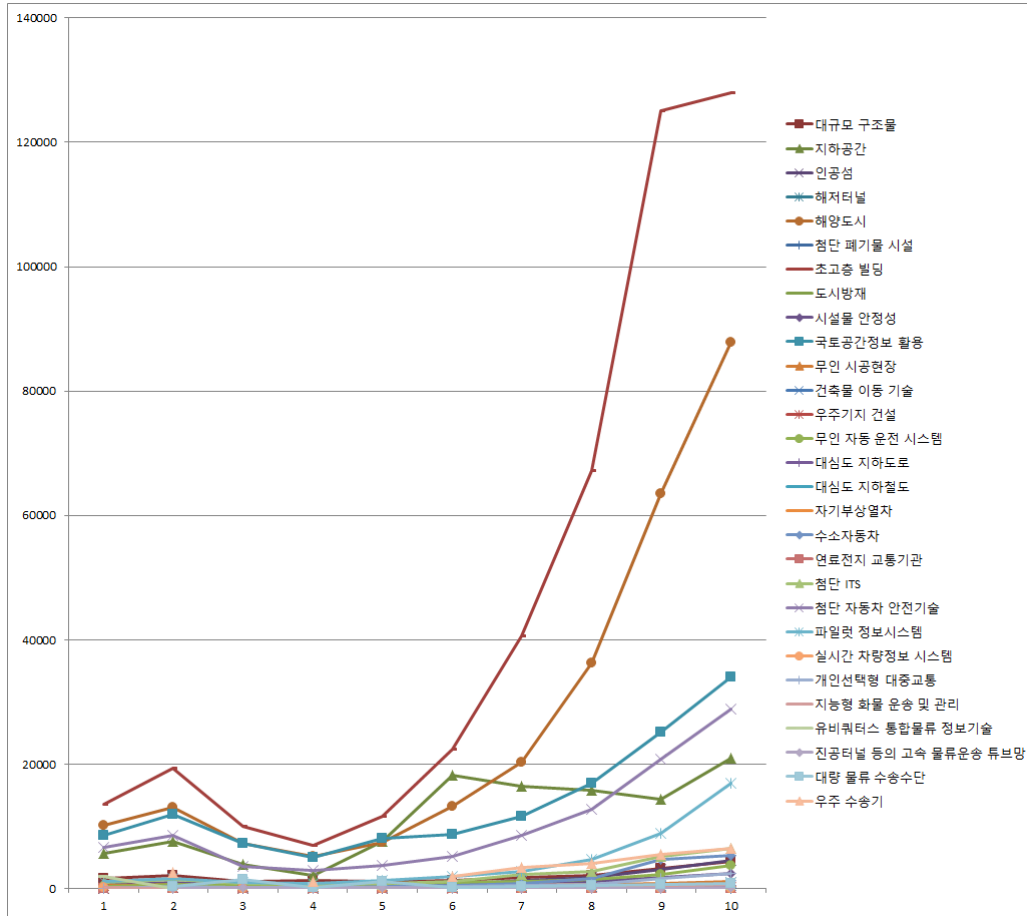
&lt;표 11&gt; 주요 기술의 우선순위에 따른 TF-DF

No	분류	기술명(한글)	기술명(영문)	TF-DF
7	건설	초고층 빌딩	Skyscraper	1.8556025
5	건설	해양도시	ocean city	1.0999907
10	건설	국토공간정보 활용	spatial information	0.6536529
2	건설	지하공간	deep underground	0.5906837
21	교통	첨단 자동차 안전기술	Transportation safety	0.439549
22	교통	파일럿 정보시스템	Head Up Display(HUD)	0.169426
29	교통	우주 수송기	space transportation	0.1127485
1	건설	대규모 구조물	large structure	0.1015096
20	교통	첨단 ITS	intelligent transportation system	0.0991385
14	교통	무인 자동 운전 시스템	unmanned vehicles	0.0594938
3	건설	인공섬	Artificial island	0.0588668
18	교통	수소자동차	hydrogen car	0.0544719
9	건설	시설물 안정성	facility safety	0.0479931
28	교통	대량 물류 수송수단	large transportation	0.0371752
11	건설	무인시공현장	automated construction	0.0271896
24	교통	개인선택형 대중교통	Personal Rapid Transit(PRT)	0.026965
4	건설	해저터널	undersea tunnel	0.0117184
17	교통	자기부상열차	magnetic levitation train	0.0109483
12	건설	건축물 이동 기술	building motion	0.0101684
25	교통	지능형 화물 운송 및 관리	intelligent logistics	0.007569
27	교통	진공터널 등의 고속 물류 운송 튜브망	tube transportation	0.0050366
13	건설	우주기지 건설	space station construction	0.0047533
15	교통	대심도 지하도로	deep road	0.004304
26	교통	유비쿼터스 통합물류 정보기술	ubiquitous logistics	0.004052
8	건설	도시방재	urban disaster prevention	0.0026285
19	교통	연료전지 교통기관	fuel cell transportation	0.0013928
6	건설	첨단 폐기물 시설	Waste Manifest System	0.0013735
23	교통	실시간 차량정보 시스템	Vehicle Information Communication System(VICS)	0.0010955
16	교통	대심도 지하철도	deep railroad	0.0005029



## ■ 정보분석 단계

정보분석단계에서는 검색된 데이터를 중심으로 연도별 그래프 변화 추이를 분석하였으며, 변화된 추이는 아래그림과 같다.

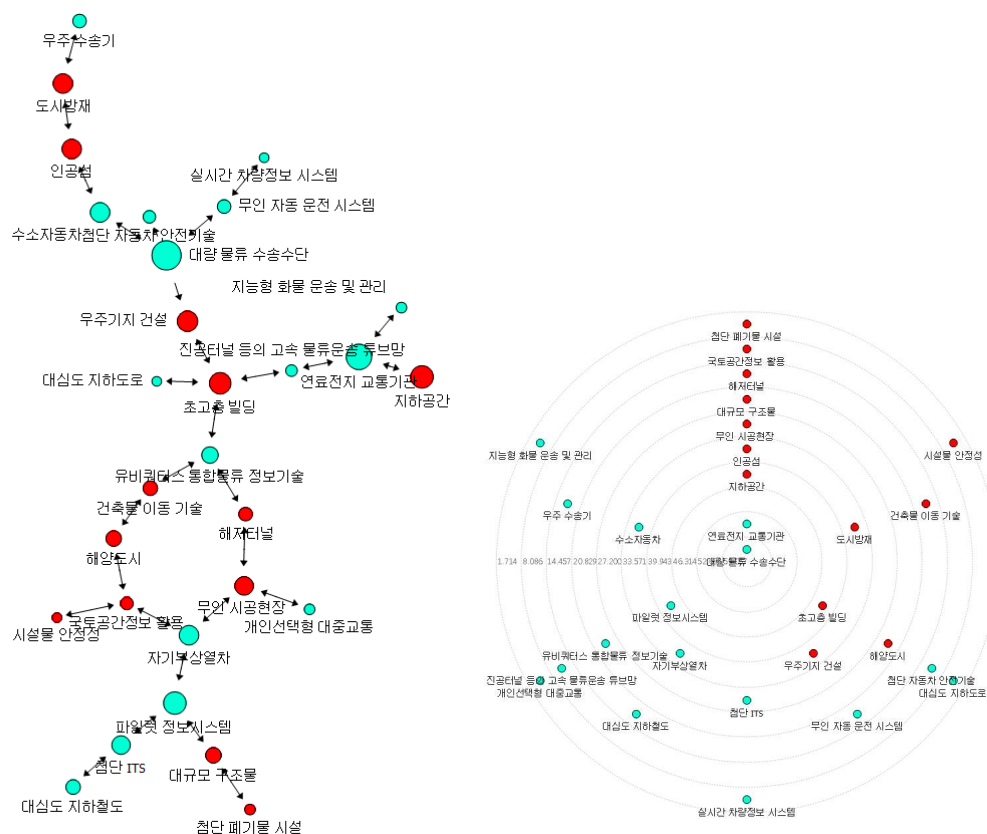


<그림 17> 주요 기술의 연도별 변화 추이

## ■ 네트워크 분석단계

미래분야의 관련된 건설분야의 네트워크 분석한 과정은 다음과 같다.

- 건설분야 29개 요소기술에 대한 노드로 선정
- Google AJAX Search API를 이용하여 요소기술간 상호 연관관계 도출(검색결과값을 이용한 가중치 선정)
- 패스파인더방식(Pathfinder : PFNet) 방식을 이용하여 비방향성의  $n(n-1)/2$  링크를 가중치의 중요도에 따라 간소화 및 최적화



### 제3절 새로운 예측방법과의 비교 분석

#### 1. 비교분석 대상

본 연구에서는 텍스트마이닝과 복잡계 네트워크를 이용하여, 건설분야를 대상으로 미래예측에 필요한 키워드를 도출하여 우선순위를 분석하였다. 특히, 건설분야의 분석대상은 ‘제3회 과학기술예측조사 수정·보완’(2008)(이하 ‘제3회 과학기술예측조사’라 함)에서 분석한 건설분야의 기술과 ‘과학기술 미래비전’(2010)을 대상으로 미래예측에 필요한 키워드들을 분석하였다. 또한, 2개의 모집단에서 분석되어진 키워드들은 정부에서 2007년에 발표된 ‘건설교통 R&D 중장기계획(2008~2012)’(2007)과 비교하여 건설분야 기술 개발 방향을 살펴보았다.

우선 시나리오에 이용한 과학기술 미래비전과 전문가 델파이에 의해 조사된 제3회 과학기술예측조사에서 건설분야의 미래를 예측할 수 있는 키워드들 도출하기 위하여 2단계 분석을 실시하였다. 1단계분석에서는 여러 가지 분석대상에서 최근의 트렌드를 반영한 키워드를 도출하기 위한 방법으로 텍스트마이닝 기법을 활용하였다. 2단계에서는 도출된 키워드들을 중심으로 키워드간의 우선순위를 분석하였다. 최종적으로는 중장기 기술을 도출한 제3회 과학기술예측조사와 시나리오 분석에 의한 중장기 과학기술 미래비전에서 도출된 건설분야 기술을 ‘건설교통 R&D 중장기계획(2008~2012)’과 비교하였다.

1단계 텍스트마이닝 분석을 위하여, 과학기술 미래비전과 제3회 과학기술예측조사에서 반복적으로 노출된 키워드들을 중심으로 새로운 키워드 그룹을 형성하고, 이 키워드들의 인터넷에 노출된 정보 검색결과량을 바탕으로 키워드의 빈도수를 분석하는 방식으로 진행하였다. 그리고 TF-DI (Term Frequency - Date Index)와 구글 API를 이용한 프로그램을 이용

하여 분석하였다. 이러한 방법들은 현재의 키워드가 인터넷에 노출된 정보의 양을 빈도수로 이용한 것으로, 여기에 최근 년도를 기준으로 가중치를 사용함으로써 트렌드 분석이 가능하도록 설계된 것이다.

2단계분석은 텍스트마이닝으로 도출된 키워드들을 대상으로 서로의 우선순위를 분석하였다. 특히, 한정된 키워드들 간의 연결고리가 있다는 가정하에 복잡계 네트워크 방법을 이용하여 분석하였다. 또한 네트워크의 중심성을 집중적으로 분석하였으며, 이러한 분석이 가능하도록 노드와 링크간 가중치 생성과 노드간의 링크의 최소화 및 최적화 방법을 사용하였다.

이러한 2단계 분석에서 사용되어진 구체적인 기술들을 살펴보면, 텍스트마이닝 기법중 하나인 TF-DI(Term Frequency - Date Index)을 개발하고 구글 AJAX Search API를 이용한 프로그램, 복잡계 네트워크의 중심성 활용, 패스파인더 방식 등 정량적으로 분석이 가능할 수 있도록 시도하였다.



<그림 19> 미래예측을 위한 키워드 분석 프로세스

## 2. 텍스트마이닝 단계에서의 분석

### ■ 제3회 과학기술예측조사

텍스트마이닝 분석단계에서는 4단[비정형 정보수집 → 정보처리 → 정보추출 → 정보분석]의 절차에 따라 분석되어진다.

비정형정보수집 과정에서는 제3회 과학기술예측조사의 기술중 건설·교통 분야의 11개의 중점과학기술에 따른 42개의 중점과학기술 세부기술 선정하였다.

정보처리 과정은 도출된 대상 자료와 기술목록을 중심으로 각 기술에 해당되는 정보를 추출할 수 있는 데이터 소스를 가공하는 단계로 11개의 중점과학기술의 개요서를 중심으로 42개의 중점과학기술 세부기술에 대한 수집된 자료(문서)를 기본으로 관련 키워드의 추출과 추출된 키워드 영문화 작업을 실시하였다.

정보추출 과정에서는 TF-DI(Term Frequency - Date Index)와 프로그램을 이용하여 미래 키워드를 예측하는 모델에 적용하고 모델적용에 따른 주요 키워드 도출하였다. 구글 검색엔진을 활용한 검색 결과값을 가중치는 2000.1.1 ~ 2009.12.31(10년간)을 대상으로 하였다.

정보분석 단계에서는 중점과학기술의 카테고리별로 그룹핑된 기존의 영문키워드와 새로 도출된 키워드를 분석하여 새로운 연관성이 있는 기술들의 그룹명과 그룹핑을 실시하였다. 또한, 최종적으로 30개의 과제를 도출하였다.

&lt;표 12&gt; 제3회 과학기술예측조사 텍스트마이닝 결과

중분류	no	기술리스트	영문키워드
건설 기반기술	1	지능형 시설 유지관리	intelligent facility maintenance
	2	시설안전관리	facility safety management
	3	에너지/자원절약형건설기술	green building
	4	첨단 건설재료	advanced construction material
	5	건설기자재 품질인증검사시험 기술	construction quality certification
	6	건설표준화	Construction Standard
	7	건설 생산시스템	construction production system
거대 구조물	8	초고층 빌딩 설비 및 환경	Skyscraper facility
	9	대심도 지하공간	deep underground
	10	초장대 교량 장비	long-span bridge equipment
	11	복합공간개발	complex spatial
미래 철도	12	초고속자기부상열차기술	magnetic levitation train
	13	400km/h 분산형 고속철도기술	high-speed rail
	14	첨단경량전철	light rail
효율적인 국토관리	15	U-GIS	Ubiquitous GIS
	16	국토 모니터링	Land Monitoring
	17	공간정보 기반 인프라	spatial information
항공 및 해양	18	항공 교통·운항 효율화	Air Transport efficiency
	19	항공안전기술	Aviation Safety
	20	해양안전기술	Marine safety
	21	해양교통관리기술	marine transportation management
미래 교통	22	ITS 기술	intelligent transportation system
	23	미래 교통안전향상 기술	Transportation safety
	24	교통 통합 시스템	Integration Transportation
미래 주거환경	25	친환경 도시건설	green city
	26	도시 재생 시스템	urban regeneration
	27	실버 주거환경	silver housing
	28	인텔리전트 교육시설	intelligent education facility
첨단 물류	29	해운물류 기술	shipping logistics
	30	물류 자동화	logistics automation

## ■ 과학기술 미래비전 분석

비정형 정보수집 과정은 2010.2월에 발표된 과학기술 미래비전의 내용 중 건설분야와 관련된 내용을 추출하여 분석하였다. 특히, 건설분야와 관련된 미래모습은 “자연과 함께하는 세상, 풍요로운 세상, 건강한 세상, 편리한 세상”의 4분야로 미래모습을 선정하였으며 그중 편리한 세상의 6대 트렌드 중 건설분야와 관련된 ‘복합공간과 생태도시 개발 기술은 육상 공간의 과밀화와 생활 패턴 변화 등에 대응하기 위해서 점차 그 중요성이 높아질 것이다’와 ‘새로운 물류·운송 수단의 등장으로 교통수단의 효율성이 증대되어 이동 시간이 단축되고 생활권이 확대될 것이다’의 시나리오와 미래상을 대상으로 분석하였다.

정보처리 과정에서는 시나리오에 의해 작성된 미래 트렌드에서 주요 키워드 도출하고, 도출된 키워드의 수정 및 검토를 통해 영문화 작업을 진행하였다.

정보추출 단계에서 도출된 영문키워드를 중심으로 TF-DF(Term Frequency - Date Frequency)를 활용하여 키워드의 가중치를 도출하였다. 구글 검색엔진을 활용하여 검색 결과값을 가중치로 도출하였으며, 검색기간은 2000.1.1 ~ 2009.12.31(10년간)을 대상으로 하였다.

정보분석 단계에서는 건설과 교통분야를 분류하여 최종적으로 총 29개의 키워드를 도출하였다.

&lt;표 13&gt; 미래비전 텍스트마이닝 결과

구분	기술명(한글)		기술명(영문)
건설	1	초고층 빌딩	Skyscraper
	2	해양도시	ocean city
	3	국토공간정보 활용	spatial information
	4	지하공간	deep underground
	5	대규모 구조물	large structure
	6	인공섬	Artificial island
	7	시설물 안정성	facility safety
	8	무인시공현장	automated construction
	9	해저터널	undersea tunnel
	10	건축물 이동 기술	building motion
	11	우주기지 건설	space station construction
	12	도시방재	urban disaster prevention
	13	첨단 폐기물 시설	Waste Manifest System
교통	14	첨단 자동차 안전기술	Transportation safety
	15	파일럿 정보시스템	Head Up Display(HUD)
	16	우주 수송기	space transportation
	17	첨단ITS	intelligent transportation system
	18	무인 자동 운전 시스템	unmanned vehicles
	19	수소자동차	hydrogen car
	20	대량 물류 수송수단	large transportation
	21	개인선택형 대중교통	Personal Rapid Transit(PRT)
	22	자기부상열차	magnetic levitation train
	23	지능형 화물 운송 및 관리	intelligent logistics
	24	진공터널 등의 고속 물류운송 튜브망	tube transportation
	25	대심도 지하도로	deep road
	26	유비쿼터스 통합물류 정보기술	ubiquitous logistics
	27	연료전지 교통기관	fuel cell transportation
	28	실시간 차량정보 시스템	Vehicle Information Communication System(VICS)
	29	대심도 지하철도	deep railroad



## ■ 텍스트마이닝단계의 비교분석

텍스트마이닝 과정을 통해 도출된 과학기술 미래비전과 제3회 과학기술 예측조사의 키워드들을 분석하여 유사한 키워드들간 그룹핑을 하였다.

그 결과, 초고층빌딩, 국토 공간이용, 지하공간, ITS, 대량물류 수송시스템, 자기부상열차, 새로운 교통수단, 교통정보시스템, 교통안전, 친환경 도시 등의 키워드들은 유사하게 나타났다.

시나리오 기법을 이용한 과학기술 미래비전은 제3회 과학기술예측조사보다 먼 미래를 예측하다보니, 우주기지 및 우주 수송선, 건축물 이동기술 등과 같은 좀더 상당히 미래지향적인 기술들이 나타났다.

반면 제3회 과학기술예측조사는 경량철도, 초고속철도, 첨단 건설재료, 건설시공 등 가까운 미래에 실현이 가능한 기술들이 포함되어 있다.

<표 14> 텍스트마이닝단계에서 도출된 키워드

과학기술 미래비전	제3회 과학기술예측조사
초고층 빌딩	초고층 빌딩 설비 및 환경
국토공간정보 활용	국토 모니터링
도시방재	공간정보 기반 인프라
지하공간	대심도 지하공간
대심도 지하도로	
대심도 지하철도	
첨단ITS	ITS기술
대량 물류 수송수단	해운물류 기술
지능형 화물 운송 및 관리	물류 자동화
진공터널 등의 고속 물류운송 튜브망	
유비쿼터스 통합물류 정보기술	
실시간 차량정보 시스템	교통 통합 시스템

과학기술 미래비전	제3회 과학기술에 측조사
자기부상열차	초고속자기부상열차기술
시설물 안정성	시설안전관리
	지능형 시설 유지관리
첨단 자동차 안전기술	미래 교통안전향상 기술
무인 자동 운전 시스템	
첨단 폐기물 시설	친환경 도시건설
	도시 재생 시스템
해양도시	복합공간개발
우주기지 건설	
인공섬	
우주 수송기	400km/h분산형고속철도기술
수소자동차	첨단경량전철
연료전지 교통기관	
개인선택형 대중교통	
파일럿 정보시스템	항공 교통·운항 효율화
	항공안전기술
	해양안전기술
	해양교통관리기술
	실버 주거환경
건축물 이동 기술	인텔리전트 교육시설
대규모 구조물	
무인시공현장	에너지/자원절약형건설기술
해저터널	첨단 건설재료
	건설기자재 품질인증·검사·시험 기술
	건설표준화
	건설 생산시스템
	초장대 교량 장비
	U-GIS

### 3. 복잡계 네트워크 분석단계

#### ■ 네트워크 분석 개요

제3회 과학기술예측조사에서 도출된 30개의 키워드들을 복잡계 네트워크로 분석하기 위하여 사전작업을 실시하였다. 특히, 구글 검색프로그램을 활용한 키워드간 연결관계에 대한 가중치를 도출하였으며, 도출된 링크의 최소화 및 최적화를 위하여 패스파인더방식을 활용하였다. 그리고 최적화된 네트워크는 3가지의 중심성 분석을 통해 우선순위를 결정하였다.

연결중심성(Degree Centrality)은 어떤 기술이 다른 기술들과 가장 많이 링크되어 있는지의 정도를 파악하는 방법으로, 허브역할을 담당하는 기술들을 분석하였다.

매개중심성(Betweenness Centrality)은 기술간의 중개역할을 담당하는 기술로 기술간의 융합기술에 해당하는 기술들을 도출하였다.

근접중심성(Closeness Centrality)은 전체 네트워크에서 어떤 기술이 중심에 있는가를 파악하는 방법으로, 네트워크 내에 가장 많은 영향력을 미치는 기술들을 도출하였다.

#### ■ 제3회 과학기술예측조사 네트워크 분석

연결중심성(Degree Centrality)을 분석한 결과 친환경 도시건설, 첨단 건설재료, 도심도 지하공간, 400 km/h 분산형 고속철도기술, 미래교통안전향상 기술 순으로 나타났다. 이러한 기술들은 다른 기술들의 허브역할을 담당하고 있으며, 가장 중요한 기술들로 나타났다. 향후 친환경적인 대단위 주거환경과 초고속 철도 등의 대중교통 수단이 기술이 중요한 기술로 나타났다.

&lt;표 15&gt; 제3회 과학기술예측조사결과의 연결중심성 분석

우선순위	중심성 분석	
	미래 기술	지수
1	친환경 도시건설	648.448276
2	첨단 건설재료	629.310345
3	대심도 지하공간	587.931034
4	400 km/h 분산형 고속철도기술	574.137931
5	미래교통안전향상 기술	541.724138

매개중심성(Betweenness Centrality) 분석에서는 친환경 도시건설, 미래 교통안전향상 기술, 첨단경량전철, 첨단 건설재료 등의 순으로 나타났다. 이러한 기술들은 기술간의 중개역할을 하는 중요한 기술로 나타났다. 친환경 도시건설과 첨단 건설재료는 허브역할을 담당하면서 타 기술간의 중요한 매개역할을 담당하고 있다.

&lt;표 16&gt; 제3회 과학기술예측조사결과의 매개중심성 분석

우선순위	매개중심성 분석	
	기술	지수
1	친환경 도시건설	0.535714
2	첨단 경량전철	0.413793
3	첨단 건설재료	0.400246
4	미래교통안전향상 기술	0.413793

근접중심성(Closeness Centrality)은 전체 네트워크에서 어떤 기술이 중심에 있는가를 파악하는 방법으로, 네트워크 내에 가장 많은 영향력을 미치는 기술로는 친환경 도시건설과 초장대 교량장비 기술이다.

&lt;표 17&gt; 제3회 과학기술예측조사결과의 근접중심성 분석

우선순위	근접중심성 분석	
	기술	지수
1	친환경 도시건설	0.273585
1	초장대 교량장비	0.273585

## ■ 과학기술 미래비전

연결중심성(Degree Centrality)을 분석한 결과, 대량 물류 수송수단, 연료전지 교통기관 순으로 나타났다. 과학기술 미래비전에서는 대량 물류수송과 새로운 교통기관과 관련된 기술들이 허브역할을 담당하고 있는 것으로 나타났다. 특히, 미래에서는 교통수단이 중

&lt;표 18&gt; 과학기술 미래비전결과의 연결중심성 분석

우선순위	중심성 분석	
	기술	지수
1	대량 물류 수송수단	65.428571
2	연료전지 교통기관	54.428571

매개중심성(Betweeness Centrality) 분석에서는 기술간의 중개역할을 담당하는 중요한 기술로써, 초고층빌딩, 유비쿼터스 통합물류 정보기술 순으로 나타났다. 이러한 결과로 인해 허브를 담당할만큼 중심적이지 않지만 타 기술간에서 연결고리 및 연관관계를 나타내는 중요한 기술이다.

&lt;표 19&gt; 과학기술 미래비전결과의 매개중심성 분석

우선순위	중심성 분석	
	기술	지수
1	초고층빌딩	0.648148
2	유비쿼터스 통합물류 정보기술	0.526455

근접중심성(Closeness Centrality)은 전체 네트워크에서 어떤 기술이 중심에 있는가를 파악하는 방법으로, 네트워크 내에 가장 중심에 있으면서 많은 영향력을 많이 미치는 기술로 나타났다. 즉, 모든 기술과 가장 근접해 있는 기술이다.

&lt;표 20&gt; 과학기술 미래비전결과의 근접중심성 분석

우선순위	중심성 분석	
	기술	지수
1	초고층빌딩	0.280000

### ■ 제3회 과학기술예측조사와 과학기술 미래비전 비교분석

제3회 과학기술예측조사에서는 친환경 도시, 도심도 지하공간 등의 인간의 풍요로운 생활환경을 위한 주거공간의 중요한 것으로 나타났으며, 고속철도 및 첨단 경량철도 등의 철도교통이 중요한 교통수단으로 나타났다.

과학기술 미래비전에서는 대량 물류 수송수단, 연료전지 교통기관 등 새로운 형태의 교통수단이 나타날 것으로 예상되며, 주거공간은 초고층빌딩이 주로 나타날 것으로 예상되었다.

제3회 과학기술예측조사와 과학기술 미래비전을 비교분석한 결과, 미래에서는 친환경적이며 거대하고 집합적인 주거환경과 대량의 물류 및 신속한 운송 등의 통철도와 유사한 형태의 대중교통수단이 중요한 기술로 등장할 것으로 예상된다.

<표 21> 예측조사와 미래비전의 비교분석 결과

중심도	제3회 과학기술예측조사		과학기술 미래비전	
	순위	기술분야	순위	기술분야
연결중심성	1	친환경 도시건설	1	대량 물류 수송수단
	2	첨단 건설재료	2	연료전지 교통기관
	3	대심도 지하공간	—	
	4	400 km/h 분산형 고속철도기술	—	
매개중심성	1	친환경 도시건설	1	초고층빌딩
	2	첨단 경량전철	2	유비쿼터스 통합물류 정보기술
근접중심성	1	친환경 도시건설	1	초고층빌딩
	1	초장대 교량장비		

#### 4. 건설분야의 비교분석

제3회 과학기술예측조사와 과학기술 미래비전은 과학기술 전문가들에 의해 작성된 중장기적 미래예측 자료로써 향후 20~30년후의 미래에 대한 키워드들과 방향을 도출한 것이다. 특히, 제3회 과학기술예측조사는 전문가 델파이에서 조사분석된 자료이며, 과학기술 미래비전은 시나리오 기법에 의해 만들어진 자료이다.

이러한 중장기적 미래방향과 현재 국가연구개발사업으로 투자되고 있는 건설교통분야의 기술들을 비교하기 위해, 현재의 건설분야 R&D 중장기 계획을 살펴보면 다음과 같다.

2007년 전문가에 의해 발표된 2008~2012년까지 국가연구개발사업의 투자를 위한 건설교통 R&D중장기계획(안)을 살펴보면, 정부주도에 의해 전략적으로 투자해야될 기술과, 단기일내에 성과를 실현해야할 기술, 미래지향적인 기술들이 혼재되고 어 존재한다.

단기적인 연구성과에 효율적인 투자를 위한 건설분야 R&D 중장기계획이지만 전체적인 흐름을 분석하면 초고층, 환경친화적인 도시, 고속철도, 지능형 도로, 재해예방 기술 등의 크게 나타나고 있다. 특히, 거대하고 집합적인 초고층건물과 이러한 건물들로 구성된 친환경 도시, 고속 및 대량 운송수단, 재해예방 기술 등을 예측해 볼 수 있다.

<표 22> 중장기계획(안)에서 우선순위가 높은 기술분야

사업	전략프로젝트
건설기술혁신사업	스마트하이웨이, 초장대 교량, 차세대 홍수 방어 등
플랜트고도화기술사업	도시기반 에너지·환경 복합플랜트 등
첨단도시개발사업	U-Eco City, 초고층 빌딩, 입체형 도시재생, 지능형 국토정보 등
교통체계효율화사업	안전지향형 교통 환경개선 기술개발, 바이모달 수송 시스템 개발, 이용자 맞춤 및 연계 환승기술, 지속가능 교통물류시스템 기술개발, 물류시설 고도화 기술개발, 물류정보화 및 자동화 기술 등
미래철도사업	차세대 고속철도시스템 기술개발, 경량전철시스템 기술개발 등
항공선진화사업	항공 사고예방 기술개발, 첨단 항공통신 시스템 기술개발 등



제3회 과학기술예측조사, 과학기술 미래비전, 건설교통 R&D중장기계획(안)을 살펴보면, 건축물의 거대화 및 집적, 도시환경 시스템이 중요성, 대량 물류 수송과 빠른 이동수단 등이 공통적으로 나타났다. 또한, 제3회 과학기술예측조사, 과학기술 미래비전에서는 홍수 등 자연재해예방기술이 따로 없었으며, 이는 미래의 주거환경에서는 자연재해로부터 오는 피해를 최소화할 수 있는 요소들이 시스템속에 기본적으로 포함될 것으로 예상된다.

건설분야는 인간의 신체 및 구조 특성상, 다른 분야의 나노, 디지털 등과 같이 변화될 소지가 적으며, 일반적인 의식주를 위해서는 아날로그적인 행동패턴과 방식들이 그대로 존재하고 있다. 따라서, 과거부터 지금까지 주거를 위한 기본적인 공간과 방법들은 현재까지 별다른 변화가 없었다. 특히, 초고층빌딩이나 대심도 지하공간은 규모상으로는 대단위로 커졌지만 형태상으로 과거부터 지금까지 유사하다.

그러나, 미래에는 건설재료의 형상화와 물리적인 변화, 도시수준의 환경 제어, 교통수단의 융합 등으로 거대한 종합시스템의 컨트롤이 필요할 것으로 예상된다. 특히, 전체 사회속에서 구성원들 각각의 개인에 대한 권리와 책임, 존엄성이 커질 것으로 예상되며, 건설분야에서도 패러다임이 바뀔 것으로 예상된다. 즉, 전체속에서의 사회구성원들의 역할이 아니라, 개인적인 사회구성원들의 수요를 만족하는 전체의 구성으로 패러다임이 바뀔 것으로 예상된다.

따라서, 제3회 과학기술예측조사, 과학기술 미래비전, 건설교통 R&D중장기계획(안)을 비교분석한 결과, 거대하고 집합적인 초고층건물과 친환경 도시는 개개인의 아날로그적인 개인주거 환경을 만족하면서 전체속의 분리된 개인의 주거환경이 아니라, 개인이 필요한 주거환경을 모여서 전체적으로 만족하는 환경으로 변할 것으로 예상된다. 이러한 패러다임의 변화는 과거 인구대비 숫적인 많은 변화가 없더라고 개인의 영유할 시스템은 과거보다

더 많은 대단위 시스템들이 필요하게 될 것이다. 또한, 대량 물류체계도 더욱더 필요하게 될 것으로 예상된다.

철도 등 대중교통시스템은 많은 인원들을 신속하게 이동시키는 수단으로 발전하면서 개개인의 공간활용도는 더욱 커질 것으로 예상된다. 예를 들면, 타인의 간섭이 없는 개인의 공간을 보유하는 대중운송 수단도 등장할 것으로 예상된다.

이러한 변화들을 쉽게 통제하고 운용될 에너지, IT 등의 거대한 융합기술들이 등장하면서 도시수준의 시스템에 대한 통제와 운용이 가능해질 것이므로 예상되므로 더욱 대형화 될 것으로 예상된다. 또한, 대형화와 집중화에 따른 패러다임의 변화도 중요한 변화의 키워드가 될 것으로 예상된다.

<표 23> 예측조사/미래비전/중장기계획(안)의 비교분석 결과

	제3회 과학기술 예측조사	과학기술 미래비전	건설교통 R&D 중장기 계획(안)
거대건 물 및 도시환 경	친환경 도시건설	초고층빌딩	초고층
	첨단 건설재료		도시기반 에너지·환경 복합플랜트
	대심도 지하공간	—	U-Eco City
초고속 및 대량 운송수 단	초장대 교량장비	연료전지 교통기관	스마트하이웨이
	400 km/h 분산형 고속철도기술		고속철도시스템
	첨단 경량전철		경량전철시스템
		유비쿼터스 통합물류 정보기술	교통물류시스템 및 고도화
		대량 물류 수송수단	
기타			자연재해 관련 기술

## 5. 비교분석 결과

미래관련 키워드의 도출방법인 텍스트마이닝의 유용성을 분석하기 위하여, 정성적인 방법(전문가에 의한 기술 도출 방법)에 의해 도출된 미래예측 키워드들과 정량적인 방법(텍스트마이닝 기법)으로 도출된 키워드들을 비교 분석을 통해 텍스트마이닝 기법의 유용성을 증명하였다.

대조군에 해당되는 정성적인 방법인 제3회 과학기술예측조사에서는 건설 및 교통관련 42개의 기술들을 기준으로 선정하였다.<sup>49)</sup> 따라서, 제3회 과학기술예측조사에 도출된 42개의 키워드들은 기존의 전문가들에 의해 도출된 기술들이다.

실험군에 해당되는 정량적인 방법으로, 시나리오 기법에 의해 작성된 과학기술 미래비전을 활용하여 29개의 기술들을 도출하였다. 이러한 정량적인 방법은 본 연구에서 개발된 텍스트마이닝 기법중에 하나인 TF-DF (Term Frequency - Date Frequency)를 활용하여 도출하였다.

대조군인 제3회 과학기술예측조사의 기술들은 2단계의 프로세스를 통해 도출되었다. 1단계에서는 제3회 과학기술예측조사(761개), Total Roadmap 국가중점육성기술(90개, 348개 세부기술), 일본의 8회 기술예측(859개), 제3기 기본계획 중점기술(273개), EU 유망기술(40개)<sup>50)</sup> 에서 도출한 182개의 미래 전략기술과 Total Roadmap 국가중점육성기술(90개, 348개 세부기술)과 관련부처의 신규 R&D 사업과제(6개, 25개 세부기술)을 바탕으로 449개의 후보 세부기술들을 도출하였다. 또한, 2단계에서는 산학연 전문가 평가에 의해 5년간 기술적 파급효과, 삶의 질 향상, 산업 고부가치화

49) 본 연구에서는 42개의 기술들을 기준으로 텍스트마이닝을 통해 30개의 기술들을 도출하였으며, 복잡계 네트워크 분석을 통해 중요 기술들의 도출하였다.

50) Emerging Science and Technology priorities in public research policies in the EU, the US and Japan, European Commission, 2006

기여도, 국가안위 및 국제사회 기여도, 시급성 등을 고려하여 449개의 후보기술 중 100개의 중점과학기술과 411개의 세부기술로 선정하였다. 100개의 중점과학기술에서는 국가차원의 전략을 확보하기 위한 중점전략기술(40개)과 중점전략기술보다는 시급성은 덜하나 국가차원에서 중장기전략을 위하여 확보해야 되는 전략기술(60개)로 구분하였다. 이 과정에서 건설·교통과 관련된 중점전략기술 3개와 전략기술 8개를 대상으로 중점과학기술 세부기술 42개를 선정하였다.

실험군인 과학기술 미래비전은 건설 및 교통분야와 관련된 미래모습인 자연과 함께하는 세상, 풍요로운 세상, 건강한 세상, 편리한 세상의 4분야의 미래모습을 선정하였으며 그중 편리한 세상의 6대 트렌드 중 건설분야의 관련된 ‘복합공간과 생태도시 개발 기술은 옥상 공간의 과밀화와 생활 패턴 변화 등에 대응하기 위해서 점차 그 중요성이 높아질 것이다’와 ‘새로운 물류·운송 수단의 등장으로 교통수단의 효율성이 증대되어 이동 시간이 단축되고 생활권이 확대될 것이다’의 시나리오와 미래상을 대상으로 분석하였다.

이러한 42개의 대조군과 29개의 실험군에 해당하는 기술들을 분석한 결과는 다음과 같다.

<표 24> 실험군과 대조군의 비교분석 결과

구분	제3회 과학기술예측조사	과학기술 미래비전
1	초고층 빌딩시스템 및 고성능 재료 기술	초고층 빌딩
2	초고속 시공 및 공사관리 기술	
3	Pilot Project 계획 및 설계기술	
4	고성능 환경 및 M&E 요소 기술	
5	건설 생산성·안전성 향상 기술	무인 시공현장
6	에너지/자원 절약형 건설기술	

구분	제3회 과학기술예측조사	과학기술 미래비전
7	첨단 건설재료/소재 개발 기술	
8	건설기자재 품질인증·검사·시험 기술	
9	신 건설생산시스템 및 건설표준화기술	
10	지능형 시설물 유지관리 기술	시설물 안정성
11	대심도 지하공간 구축·활용기술	지하공간 대심도 지하도로 대심도 지하철도
12	초장대 교량 설계기준 및 설계기법	대규모 구조물
13	내진/내풍에 의한 거동예측 및 진동저감 기술	
14	첨단 시공기술 및 특수 장비 개발기술	
15	초고성능 재료 개발 기술	
16	U-Eco City (친환경 도시건설 기술)	해양도시 도시방재
17	복합공간 개발 기술	인공섬
18	도시 재생 시스템 기술	첨단 폐기물 시설
19	국토 모니터링 기술	
20	공간정보 기반 인프라 기술	국토공간정보 활용
21	U-GIS 기반 건설정보화 기술	
22	인텔리전트 교육시설건설기술	
23	친실버 및 친환경 주거환경 구현 기술	
24	ITS 기술	첨단 ITS
25	U-Transportation 기술	무인 자동 운전 시스템 첨단 자동차 안전기술
26	미래 교통안전향상 기술	파일럿 정보시스템
27	교통연계 및 환승시스템 기술	
28	중소형항공기 인증기술	
29	항공 교통·운항 효율화 기술	
30	항공안전기술	
31	해양안전기술	
32	해양교통관리기술	

구분	제3회 과학기술예측조사	과학기술 미래비전
33	400km/h 분산형 고속철도기술	
34	초고속자기부상열차기술	
35	첨단경량전철 및 신교통시스템기술	
36	도시형 자기부상 시스템 기술	자기부상열차
37	통합 지능형 컨테이너 시스템 기술	지능형 화물 운송 및 관리
38	u-통관 시스템기술	유비쿼터스 통합물류 정보기술
39	지능형 교통물류 정보화 기술	
40	지능형 물류센터 및 고효율장비 기술	
41	미래형 수송시스템 기술	진공터널 등의 고속 물류 운송 튜브망
42	해운물류기술	대량 물류 운송수단
기타기술		건축물 이동기술 우주기지 건설 우주수송기 수소자동차 연료전지 교통기관 개인선택형 대중교통

대조군인 제3회 과학기술예측조사의 기술과 실험군인 과학기술 미래비전을 분석한 결과, 29개의 기술중 건축물 이동기술, 우주기지 건설, 우주수송기, 수소자동차, 연료전지 교통기관, 개인선택형 대중교통 등 5개의 기술을 제외하고 24개의 기술들이 관련되어 있다. 이는 약 82%의 기술들이 일치하는 결과를 나타내고 있다.

특히, 제3회 과학기술예측조사와 비교하여 제외된 과학기술 미래비전의 기술들은(약 18% 해당) 우주기지, 우주수송기, 개인선택형 대중교통, 수소자동차, 연료전지 교통기관 등은 오히려 제3회 과학기술예측조사의 기술들보다 훨씬 더 미래지향적인 기술들이다. 또한, 과학기술 미래비전에 포함되지 않는 제3회 과학기술예측조사의 기술들은 어떤 면에서는 미래의 기술보다는 현재 및 가까운 미래의 기술들도 상당수 많이 존재하고 있다.

이러한 이유중의 하나로써, 제3회 과학기술예측조사는 과학기술 미래비전보다 훨씬 이전에 발행된 예측된 자료이다. 또한, 과학기술 미래비전은 시나리오에 의해 작성된 과학기술로써 기술적인 내용보다 장기적인 미래의 모습을 담았기 때문에 더 미래 지향적인 기술내용을 포함하고 있다.

따라서, 정성적인 방법에 의해 작성된 제3회 과학기술예측조사(대조군)의 기술들과 정량적인 방법(텍스트마이닝)에 의한 과학기술 미래비전(실험군)의 기술들을 비교분석한 결과, 정량적인 방법(텍스트마이닝)에 의해 도출된 기술들은 전문가의 정성적인 방법에 의한 도출된 기술들과 별반 차이가 없는 것으로 나타났다.

그러므로 본 연구에서 사용된 텍스트마이닝 기법은 미래예측 키워드를 도출하는 정량적인 방법으로 유용할 것으로 증명되었으며, 전문가들의 정성적인 자료를 보조하는 정량적인 자료분석 방법으로 유용할 것으로 예상된다.

또한, 텍스트마이닝 대상 자료의 수준과 종류에 따라 오히려 더 좋은 결과를 도출하는 하나의 방법으로 더 유용할 것으로 예상된다.

## 제5장 결론 및 시사점

본 연구에서는 미래예측시 전문가에 의한 정성적인 방법을 보조 및 보완하거나, 정성적인 방법을 일부 대체하는 하나의 정량적인 방법을 제시하고, 이를 건설분야에 적용하여 검증하였다.

정량적인 방법을 통해 미래관련 키워드를 도출하기 위하여 텍스트마이닝 기법과 도출된 키워드들간의 중심기술을 분석하기 위한 복잡계 네트워크 기법을 활용하였다. 즉, 본 연구에서 개발된 텍스트마이닝의 방법인 TF-DF(Term Frequency - Date Frequency) 기법을 통해 방대한 관련자료에서 현재의 트렌드를 반영한 중요한 키워드들을 도출하였다. 또한, 중요한 키워드들을 중심으로 어떤 키워드가 더 중요한지 복잡계 네트워크의 중심성 분석을 통해 도출하였다.



&lt;표 25&gt; 텍스트마이닝과 네트워크 분석의 복합기법 검증결과

미래예측 기술	텍스트마이닝 기법	복잡계 네트워크 기법
목적	<ul style="list-style-type: none"> <li>방대한 자료로부터 현재의 트렌드를 반영한 키워드(기술) 도출</li> </ul>	<ul style="list-style-type: none"> <li>텍스트마이닝으로 도출된 키워드(기술)들을 네트워크 분석을 통한 중심 키워드(기술) 도출</li> </ul>
기술 및 알고리즘	<ul style="list-style-type: none"> <li>TF-DF(Term Frequency - Date Frequency)</li> <li>Google AJAX Search API를 활용한 인터넷 검색결과 분석 프로그램</li> </ul>	<ul style="list-style-type: none"> <li>복잡계 네트워크 중심성 분석</li> <li>Pathfinder를 이용한 최적화</li> <li>Google AJAX Search API를 활용한 기술간 상호연관관계 검색결과 분석 프로그램</li> </ul>
검증방법	<ul style="list-style-type: none"> <li>대조군 : 제3회 과학기술예측조사에서 전문가들에 의해 도출된 건설 및 교통분야 기술(42개 기술)</li> <li>실험군 : TF-DF기술을 이용하여 과학기술 미래비전에서 건설 및 교통분야 기술(29개 기술)</li> </ul>	<ul style="list-style-type: none"> <li>대조군 : 제3회 과학기술예측조사에서 도출된 중심성 기술</li> <li>실험군 : 과학기술 미래비전에서 도출된 중심성 기술</li> </ul>
결과	<ul style="list-style-type: none"> <li>실험군 23개(약 79%)의 기술들이 유사한 것으로 나타났음</li> <li>실험군에서 일치하지 않는 6개의 기술(21%)들은 대조군보다 훨씬 더 미래지향적인 기술들로 나타남</li> </ul>	<ul style="list-style-type: none"> <li>제3회 과학기술예측조사 : 친환경 도시건설, 첨단 건설재료, 고속철도 및 경량철도 등 중심키워드 도출</li> <li>과학기술 미래비전 : 대량물류수단, 연료전지 교통기관, 초고층빌딩 등 중심 키워드 도출</li> </ul>

텍스트마이닝의 방법인 TF-DF(Term Frequency - Date Frequency) 기법을 검증을 위해 전문가에 의해 도출된 기술들과 텍스트마이닝 기법에 의해 도출된 기술을 비교하였다. 즉, 전문가에 의해 도출된 제3회 과학기술예측 중에서 건설 및 교통 분야와 관련된 대조군에 해당하는 42개의 기술들과 본 연구에서 개발된 TF-DF(Term Frequency - Date Frequency) 기법을 활용하여 과학기술 미래비전의 자료에서 실험군에 해당하는 29개의 중요기술을 도출하여 비교분석하였다.

이러한 기술들을 비교분석한 결과, 실험군의 29개의 기술 중 23개의 기술들이 대조군의 기술들과 일치하였으며(약 79%), 6개의 기술들(우주기지, 우주수송기, 개인선택형 대중교통, 수소자동차, 연료전지 교통기관)은 제3

회 과학기술예측에서는 없는 더 미래지향적인 기술들로 나타났다.

따라서, 미래의 키워드들을 도출하기 위한 정량적인 방법으로 활용하는 무리가 없을 것으로 예상된다. 또한, TF-DF(Term Frequency - Date Frequency)기법에서는 제공되는 자료에 따라 유용한 키워드들을 도출할 수 있는 방법으로 더욱 유용할 것으로 예상된다.

이렇게 전문가에 의해 도출된 제3회 과학기술예측조사의 키워드와 텍스트마이닝 방법에 의해 도출된 키워드를 분석한 결과, 제3회 과학기술예측 조사에서는 친환경 도시, 대심도 지하공간 등의 도시규모의 새로운 주거공간의 중요한 것으로 나타났으며, 고속철도 및 첨단 경량철도 등의 철도교통이 중요한 대중 교통수단으로 나타났다. 또한, 과학기술 미래비전에서는 대량 물류 수송수단, 연료전지 교통기관 등 새로운 형태의 교통수단이 나타날 것으로 예상되며, 생활 주거공간과 건축기술의 패러다임을 바꾸는 초고층빌딩이 나타날 것으로 예상된다.

따라서, 건설 및 교통분야의 미래 트렌드를 간단히 분석하면, 건설분야에서는 대규모적인 주거공간과 그로 인한 생활패턴 및 방식의 패러다임이 바뀔 것으로 예상되고, 교통분야에서는 빠르고 대량의 물류 및 사람의 수송수단과 새로운 현재의 개인 교통수단을 대체하는 새로운 방식의 교통수단이 나타날 것으로 예상된다.

이와 같은 연구결과를 통해 본 연구에서 제안하는 텍스트마이닝 방법을 활용한 미래 키워드를 도출하는 방법은 매우 유용할 것으로 예상되며, 더욱 다양한 방법으로 적용이 가능할 것으로 예상된다. 또한, 복잡계 네트워크를 활용한 분석방법은 미래의 트렌드를 분석하는 하나의 방안으로 유용하게 활용될 것으로 예상된다. 이러한 두가지 기법들에 트렌드를 반영할 수 있는 방법으로 검색엔진을 활용한 인터넷 활용방법은 본 연구에서도 유

용하게 활용되었으며, 앞으로도 관련된 연구에도 매우 유용할 것으로 예상된다. 따라서, 향후 본 연구와 관련하여 정량적인 분야에서 많은 연구방법론들이 개발되고 많은 검증들이 필요할 것으로 예상된다.

본 연구와 관련하여 미래를 예측할 수 있는 정량적인 방법을 제시하는 방법론과 그에 따른 시사점을 살펴보면 다음과 같다.

대부분의 미래를 예측하는 방법들은 전문가들의 정성적인 의견과 평가에 의해 이루어지고 있으며, 전문가들의 경향, 정치적 요소, 인맥 관계, 사회적 입지 등의 다양한 요소에 의해 일부 편향적인 의견이나 주장이 많이 반영되고 있는 실정이다. 또한, 정량적인 방법론들에 대한 연구와 투자는 아직까지 초기단계에 머무르고 있는 실정이다(추세의삼법, AHP 등). 따라서, 과학기술분야의 미래 예측시 전문가의 객관적인 의견과 평가가 이루어지도록 좀 더 구체적이고 객관적인 데이터와 자료를 제공하는 방법이 중요한 이슈로 떠오르고 있으며, 특히 기존의 논문, 특허 등 다양한 소스를 어떻게 객관화할 것인지에 대한 방법과 연구가 필요하다.

인터넷을 활용한 예측 기법은 인터넷의 복잡하고 다양한 정보를 활용하여 현재의 트렌드를 반영할 수 있는 유용한 수단이며, 최근 페이스북, 트위터 등 소셜네트워크(Social Network)를 활용한 인터넷은 스마트폰 등의 하드웨어적인 요소와 인간의 아날로그적인 감성을 활용한 수단으로 엄청나게 빠른 속도로 진화하고 있다. 따라서, 본 연구에서 활용한 인터넷의 검색결과를 활용한 방법과 트렌드를 분석할 수 있는 텍스트마이닝 방법에 대한 다양한 연구가 필요하다. 또한, 본 연구에서는 복잡계 네트워크의 방법론을 도입하여 활용하였지만 좀더 확장해 본다면 복잡계의 다양한 방법들은 현재에도 진화하고 있기 때문에 어떻게 활용할 것인가에 대한 지속적인 연구가 필요할 것으로 예상된다.

# <참 고 문 헌>

- 1) 안승구, Analysis on the national R&D investment activities of Major countries in 2009(KISTEP, 2009)
- 2) Michael Marien, Future Studies in the 21st Century: A Reality -Based. View, Futures, Vol. 34(2002) 261 - 281
- 3) Governance in the 21st Century: FUTURE STUDIES(OECD, Paris, 2001)
- 4) Mikko Syrjänen, Yuko Ito, Eija Ahola (editors), Foresight for Our Future Society: Cooperative project between NISTEP(Japan) and Tekes(Finland)(Tekes & NISTEP, 2009)
- 5) Cornelia Daheim, Regional Foresight in Europe - 2 Examples: Duesseldorf and Linz, WFS Conference Minneapolis(2007)
- 6) Pirjo Stähle (ed.), Five Steps for Finland's Future(Tekes, Helsinki, 2007)
- 7) Jeremy Ginsberg, Matthew H. Mohebbi, Rajan S. Patel, Lynnette Brammer, and Mark S. Smolinski & Larry Brilliant, Detecting influenza epidemics using search engine query data, Nature 457(2009) 1012-1014
- 8) Sang Hoon Lee, Pan-Jun Kim, Yong-Yeol Ahn, and Hawoong Jeong, Googling hidden interactions: Web search engine based weighted network construction, PNAS(2008)
- 9) R. Levin, D. Rubin, and J. Stinson, Forecasting. in Quantitative approaches to management, Chapter 3(1986, NY, McGraw-Hill)
- 10) James A. Dator, Advancing Futures: Futures Studies in Higher Education(Praeger, Westport, 2002)

- 11) 박병원, 과학기술예측조사를 위한 방법론 및 프레임워크 개선연구 (KISTEP, 2007)
- 12) Hariolf Grupp and Harold A. Linstone, National Technology Foresight Activities Around the Globe: Resurrection and New Paradigms, Technological Forecasting and Social Change, Volume 60, Issue 1, 2 January(1999) 85-94.
- 13) Martin Hilbert, Ian Miles, and Julia Othmer, Foresight tools for participative policy-making in inter-governmental processes in developing countries: Lessons learned from the eLAC Policy Priorities Delphi, Technological Forecasting and Social Change, Volume 76, Issue 7(2009) 880-896
- 14) 손석호, Evaluation of Korean Technology Foresight Program (KISTEP, 2008)
- 15) Schoemaker, J.H. Pau, Scenario Planning: A Tool for Strategic Thinking, Sloan Management Review. Winter(1995) 25-40
- 16) Mats Lindgren, Hans Bandhold, Scenario planning: The link between future and strategy(Macmillan, 2002)
- 17) P. Erdős and A. Réney, Publ. Math. Inst. Hung. Acad. Sci. 5, 17(1960); Bull. Inst. Int. Stat. 38, 343(1961)
- 18) 윤영수, 채승병, 복잡계 개론(삼성경제연구소, 2005)
- 19) 손동원, 사회 네트워크 분석(경문사, 2002)
- 20) 김용학, 사회연결망 이론(박영사, 2004)
- 21) 강병남, 복잡계 네트워크 과학(집문당, 2009)
- 22) B. Bollobas, Discrete Math. 33(1981), 1
- 23) D.J Watts and S.H. Strogatz, Collective dynamics of 'small-world' networks, Nature 393(1998)
- 24) R. Albert, H. Jeong, and A.L. Barabasi, Diameter of the World

- Wide Web, Nature 401(1999)
- 25) S. A. kauffman, The origin of order : Self-organization and selection in evolution(Oxford University Press, New York, Oxford)
- 26) H. Jeong, B. Tombor, R. Albert, Z.N. Oltvai, and A.L. Barabasi, Nature 407(2000)
- 27) A. Wagner and D.A. Fell, Proc. R. Soc. London B 268, 1803 (2001)
- 28) M.E. J. Newman, SIAM Review 45(2003), 167
- 29) S. Redner, Eur. Phys. J. B 4(1998)
- 30) C. Tsallis and M.P. de Albuquerque, Eur. Phys. J. B13(2000)
- 31) Rosario N. Mantegna and H. Eugene Stanley, An Introduction to Econophysics: Correlations and Complexity in Finance(Cambridge University Press, Cambridge, 1999)
- 32) M. Faloutsos, P. Faloutsos, and C. Faloutsos, Comput. Commun. Rev. 29, 251(1999)
- 33) Liljeros et al, Nature(2001)
- 34) <http://www.santafe.edu/>
- 35) 윤영수, 채승병, 복잡계 개론(삼성경제연구소, 2005)
- 36) 위키피아 <http://www.wikipedia.org>
- 37) James T. Cushing. (1998), 송진웅 옮김, 물리학의 역사와 철학 (Philosophical Concepts in Physics), 북스힐(2006)
- 38) 박장호, 유전자 결정론(환원주의적 시각)에 대한 비판적 고찰
- 39) 이명헌 경영스쿨 [http://www.emh.co.kr/xhtml/small\\_world\\_effect.html](http://www.emh.co.kr/xhtml/small_world_effect.html)
- 40) G. Sabidussi, The centrality index of a graph. Psychometrika, 31(4)(1966) 581-603
- 41) Jae-Yun Lee, A Study on the Network Generation Methods for Examining the Intellectual Structure of Knowledge Domains, Korean Society for Library and Information Science 40(2006)

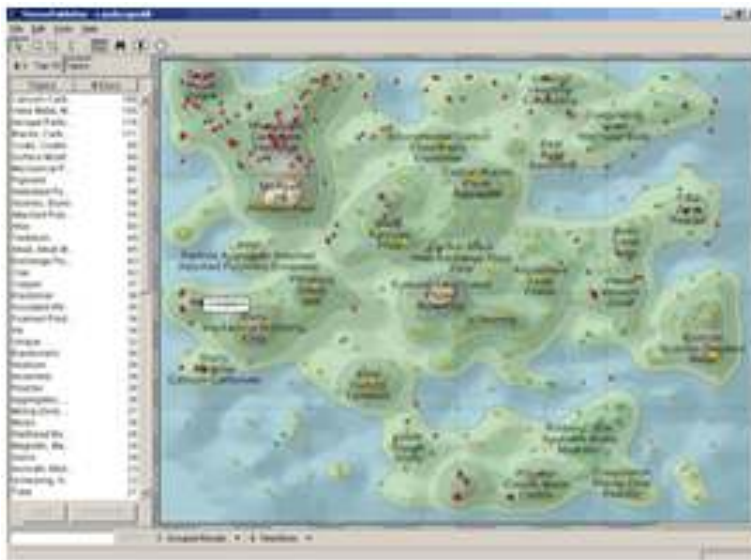
333–355

- 42) R. W. Schvaneveldt, F. T. Durso, and D. W. Dearholt, Network structures in proximity data. In G. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory*, Vol. 24(Academic Press, New York, 1989)
- 43) Quirin, A; Córdón, O; Santamaría, J; Vargas–Quesada, B; Moya–Anegón, F, A new variant of the Pathfinder algorithm to generate large visual science maps in cubic time, *Information Processing and Management*, 44, p.1611–1623(2008)
- 44) <http://www.wikipedia.org>
- 45) Salton G. and McGill, M. J. 1983 *Introduction to modern information retrieval*. McGraw–Hill
- 46) Salton G. and McGill, M. J. 1983 *Introduction to modern information retrieval*. McGraw–Hill
- 47) *Emerging Science and Technology priorities in public research policies in the EU, the US and Japan*, European Commission, 2006

## [별첨1] 텍스트마이닝 관련 소프트웨어

### ○ Aureka (<http://scientific.thomson.com/products/aureka/>)

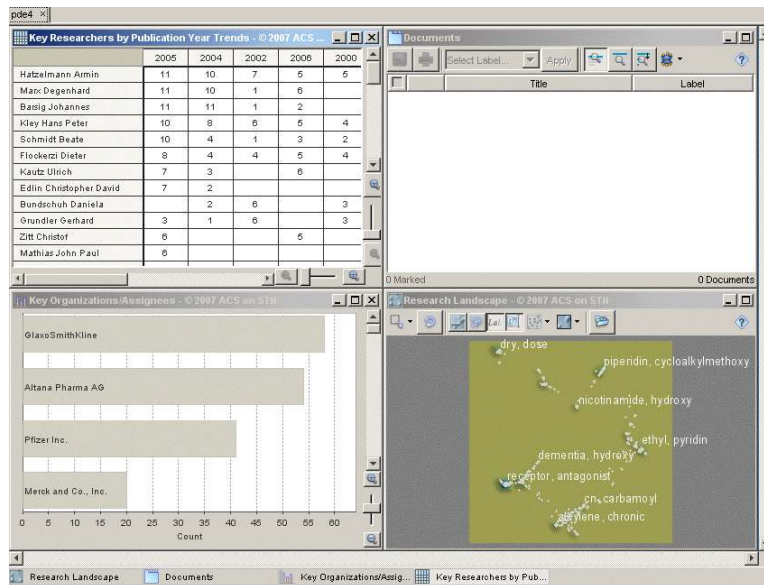
- 텍스트 마이닝 결과를 ThemeScape(R) 맵이라 불리는 등고선 지도로 나타내는 것이 특징. 이외의 대표적인 기능으로서는 인용 트리(인용·피인용 관계에 있는 특허 공보를 트리상에 표시)가 있음



### ○ STN Anavista(<http://www.cas.org/products/anavist/index.html>)

- STN 검색결과를 분석하는 소프트웨어. 다른 텍스트 마이닝 소프트웨어와는 달리, AnaVist의 소프트웨어 자체는 무료이며 해석 대상 건수에 의해 과금 체계가 다르다(비주얼화 실행료). AnaVist의 텍스트마이닝 결과는 리서치·풍경으로 불림(Sandia National Laboratories가 개발한 비주얼화 소프트웨어의 라이선스를 취득).

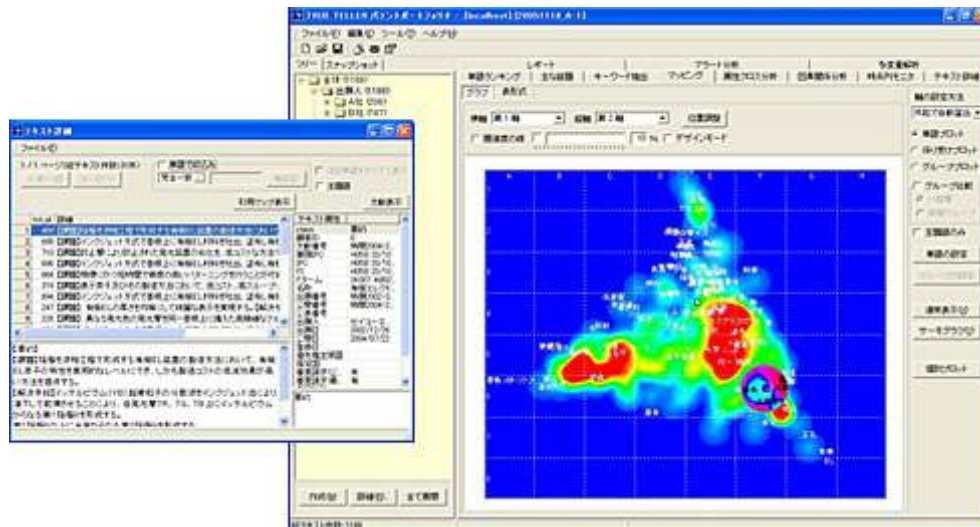




# ○ NRI : TRUE TELLER 패턴드 포트폴리오

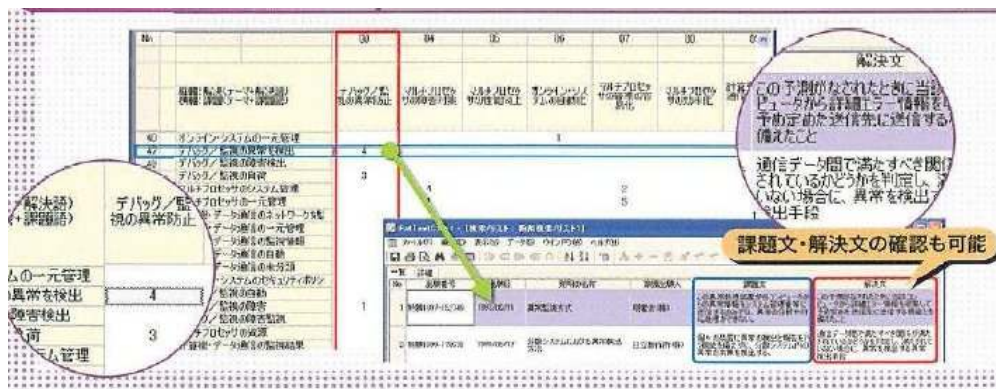
<http://www.trueteller.net/textmining/patent/>

- TRUE TELLER는 노무라 종합연구소가 개발한 텍스트마이닝툴로서, 원래는 콜 센터(고객 창구)나 앙케이트 분석 등의 텍스트 데이터를 해석하기 위한 툴을 특허 분석에 적용한 것이다. 텍스트마이닝 결과는 등고선도로 표현되는 테크놀로지 히트 맵으로 불리기도 함



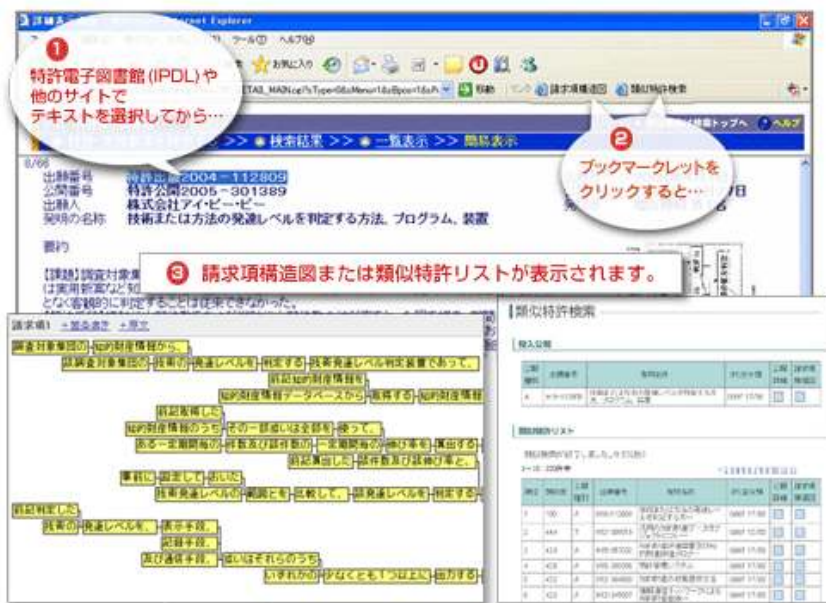
○ NEC PatTextChart(R)(<http://www.nec-nis.co.jp/product/pat/>)

- 해석 대상 특허의 명세서를 해석하고, 각 특허의 기술적 과제·해결 수단을 단문으로 자동 추출하여 맵화할 수 있는 툴. 「과제×해결책 맵」 뿐만이 아니라, 「과제×출원일 맵」 이나 「과제×출원인 맵」 도 작성 가능

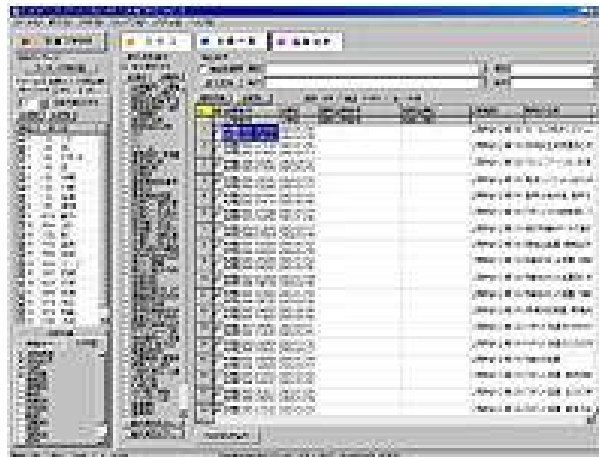


○ IPB : 패턴트 아틀라스(<https://atlas.ipb.co.jp/>)

- IPB가 제공하고 있는 서비스이며, 독자 개발한 문서 벡터 해석 시스템에 의해 자동적으로 선행 기술 조사(유사 특허 추출)·클러스터링이 가능. 클러스터링 된 결과는 특허 구조도로써 표현됨.

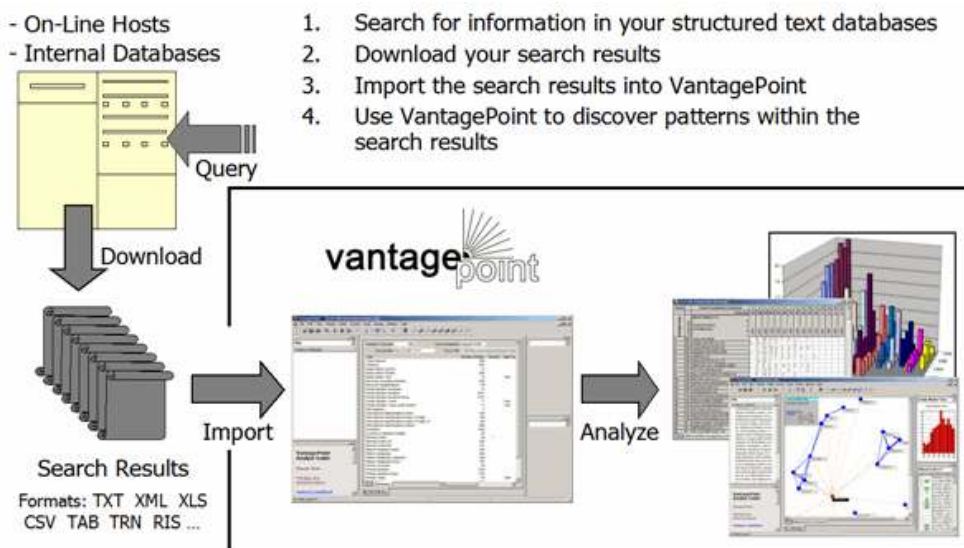
○ JPDS([http://www.jpds.co.jp/business/cd-soft\\_5.html](http://www.jpds.co.jp/business/cd-soft_5.html))

- 특허 공보의 검색 결과를 바탕으로 문서 해석·맵 작성을 실시하는 소프트웨어. 텍스트 형식의 공보 데이터의 혼잡로부터 문장 해석을 실시한다. 고도의 텍스트마이닝을 쉬운 인터페이스로 실현한 해석 툴 입문용의 소프트웨어임



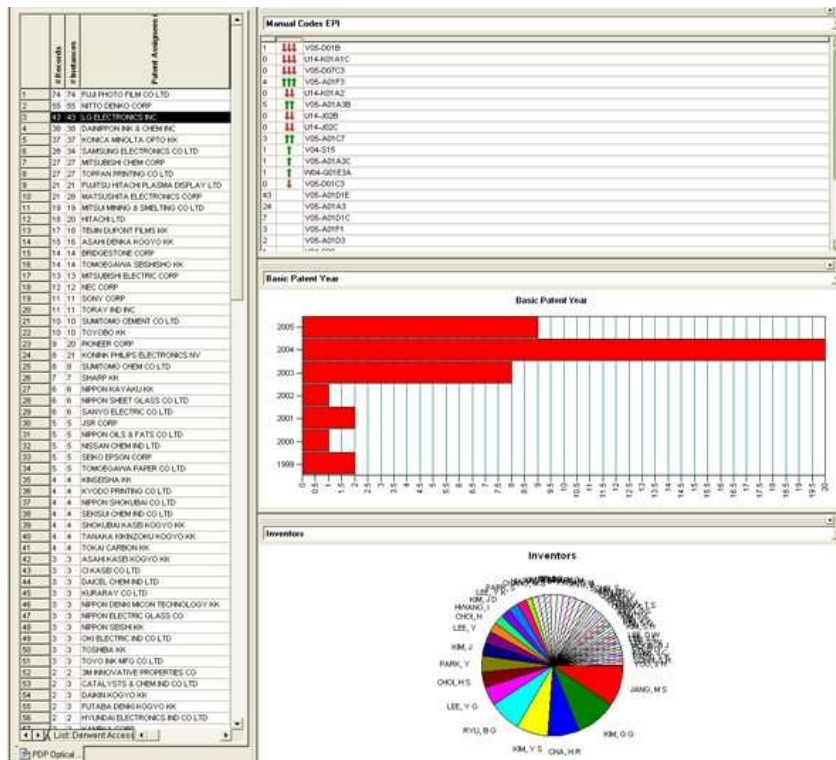
○ VantagePoint(<http://www.thevantagepoint.com/vantagepoint.cfm>)

- 미국 조지아공대 기술정책평가센터 TPAC와 Search Technology사가 공동으로 개발한 상용 분석소프트웨어. 분석대상데이터의 입력, 관리, 정제 영역에서 특히 뛰어난 기능을 갖추고 있음.



○ Thomson Data Analyzer(<http://scientific.thomson.com/products/tda/>)

- Thomson Scientific 에서 제공하는 특허분석 소프트웨어. Delphion, Dialog, STN 등의 데이터로 분석작업을 수행할 수 있음



○ 국민대 강승식교수의 한국어 형태소 분석기와 한국어 분석 모듈 (<http://nlp.kookmin.ac.kr/>)

- 현재 비상용버전으로 데이터마이닝의 핵심이 되는 한국어와 영어의 자연어분석 툴을 제시하고 있음
- 현재까지 모듈을 이용한 텍스트마이닝은 개발단계로 추정됨



○ 뉴욕주에 위치한 IBM 연구소에서 집중 연구 중인 분야 중에는 위마(UIMA)라는 정보 분석 프로젝트

- UIMA(Unstructured Information Management Architecture)는 데이터베이스처럼 정형화된 자료가 아닌 문서 자료나 음성 파일 등 비정형 정보를 분석하는 종합 시스템을 설계하는 것이 목적임
- 오픈소스 형태로 무료 배포하고 있으며, 자바 SDK 배포본 1.5 버전이 최근 공개됐다(<http://www.alphaworks.ibm.com/tech/uima>)
- UIMA는 AE(Analysis Engines)라는 블록이 여럿 결합되어 있으며, 강조하는 것은 "의미론 검색"이다. 일반적인 검색 엔진의 단순한 키워드 검색 방식보다 진일보한 검색을 추한다는 것이다. 예를 들어 “센터”라는 단어가 명칭에 들어 있는 기관을 검색하고 싶다고 생각해보자. 키워드 방식으로 “센터”라는 단어가 든 문서가 여럿 나오겠지만, 의미론 검색에서는 XML을 이용하여 “<기관>센터”라는 식으로 정확한 검색 내용을 편리하게 지정할 수 있음
- 프로젝트는 데이빗 페루치(David Ferrucci) 박사가 이끌고 있으며, 문서 분석, 기계 번역, 생물정보공학, 지식 통합, 사회적 네트워크 분석, 프로그램 분석, 시맨틱 웹, 질의응답 시스템 등 자연언어 처리나 기계학습, 정보 검색 분야의 여러 연구 결과를 응용하고 있음
- 지난 5월에 미국 동부 매클린에서 열린 지능형 시스템 학술회의(<https://analysis.mitre.org/proceedings/>)에서는 페루치와 다른 연구원들이 공저한 SAW(Semantic Analysis Workbench)에 대한 논문이 발표됐다. SAW는 일종의 소프트웨어 도구로 위마 SDK를 이용해 개발된 시스템이다.

- SAW 외에도 메드테스(MedTAS, Medical Text Analysis System, [http://www.research.ibm.com/UIMA/Project\\_MedTAS.htm](http://www.research.ibm.com/UIMA/Project_MedTAS.htm))나 BoBI(Back-of-the-Book-Indexer, [http://www.research.ibm.com/UIMA/Project\\_BoBI.htm](http://www.research.ibm.com/UIMA/Project_BoBI.htm)) 등 여러 응용 시스템이 위마 프레임워크를 기반으로 개발되어 있다. 메드테스는 의학 문서를 분석하는 도구이며, BoBI는 문서에서 색인을 뽑는 시스템이다.
- 서부에 위치한 IBM 연구소의 웹파운틴(WebFountain, <http://www.almaden.ibm.com/webfountain/>)과도 연계 되어 있으며, 웹파운틴은 비정형 정보 가운데서도 웹 정보 분석에 집중하고 있음.

## [별첨2] 중점과학기술 개요

중점과학기술 번호 87

기술 분야	건설·교통·안전 분야
기술명	초고층빌딩 건축기술
기술 정의	<p>○ 초고층 복합빌딩 건축 및 건축물의 안전성, 경제성, 효율성을 제고를 위한 초고층 빌딩의 건축 계획·구조·설비·시공분야 통합설계 및 실용화 연구 관련 기술</p>
세부기술 정의 및 범위	
<p>○ 초고층 건축물 구조시스템 및 고성능 재료 기술</p> <ul style="list-style-type: none"> <li>- 해당 건축 환경에 적합한 초고층 건축물 구조시스템 해석 및 체계적 세부 구조설계 및 최적화 관련 기술</li> <li>- 초고층 건축물에 요구되는 콘크리트 타설 기술, 최적화된 거푸집 개발 등 고성능 재료 기술</li> </ul> <p>○ 초고속 시공 및 공사관리 기술</p> <ul style="list-style-type: none"> <li>- 초고층 빌딩 건축에 요구되는 생산성 향상 및 공기지연 방지를 위해 현장 생산성 향상 공법, 가설공사 자재·시스템 개발, 발전된 공정관리기법 도입 및 새로운 자원관리 관련 기술</li> </ul> <p>○ Pilot Project 계획 및 설계기술</p> <ul style="list-style-type: none"> <li>- 향후 건축될 초고층 빌딩의 거주성 평가모델 구축을 비롯한 공간이용 효율성, 통행량, 방재 성능 예측, 빌딩 서비스 성능 설계 등 계획·설계 기술</li> </ul> <p>○ 고성능 환경 및 M&amp;E(기계·전기설비) 요소 기술</p> <ul style="list-style-type: none"> <li>- 고층부의 외부기상 조건에서 건축물 보호 및 구조적 경량화에 따른 건축 환경 문제 해결 기술 (주거건물 초고층화에 따른 난방시스템 경량화, 환기시스템, 에너지 절감기술, 정확한 부하 산정 및 M&amp;E 설비 설계 및 최적화 시스템 구축 관련 요소 기술 등 포함)</li> </ul>	



## 중점과학기술 번호 88

기술 분야	건설·교통·안전 분야
기술명	건설기반기술
기술 정의	<p>○ 미래지향적인 국토 및 도시공간의 개발·활용을 목표로 건설 프로세스 혁신 및 건설 생산성 향상과 관련된 건설기술 분야 전반의 기반기술 (첨단자재개발, 고성능·경량화, 자동화·지능화 및 IT기반 건설 프로세스 혁신, 표준화 및 인증·시험, 안전관리 네트워크 기술 등 다양한 영역 포함)</p>
세부기술 정의 및 범위	
<p>○ <b>건설 생산성·안전성 향상 기술</b> - 지능형 토공, 모듈화·사전생산-조립 등 다양한 생산성 향상기술 분야와 시설물 안전 관리 업무프로세스 개선, 수도·철도·도로 등 시설물 IT기반 안전관리 네트워크 등 안전성 향상 기술</p> <p>○ <b>에너지/자원 절약형 건설기술</b> - 건축물을 장기간 사용할 수 있게 건설함으로써 자원사용을 억제하고 사용하는 자원을 효율적으로 활용함과 동시에 재사용할 수 있도록 하는 관련 건설 기술 (에너지·자원 절약형 건축기술, 자연친화형 저가 중수처리 설비기술 외)</p> <p>○ <b>첨단 건설재료/소재 개발 기술</b> - 기존 시설물에 대한 과학적·합리적 유지관리 및 새로운 형식, 구조, 시공법이 적용된 新건축 구조물에서 요구되는 고성능의 건설자재 및 보수 및 보강용 건설재료 개발 및 활용기술</p> <p>○ <b>건설기자재 품질인증검사시험 기술</b> - 건설 자재·부품의 품질성능 확보 요구의 증대에 따라 건축물 및 토목구조물의 내구성 등 품질확보를 유도하고, 자재·부품의 품질인증, 인정, 지정, 검사 및 시험을 위한 관련 기술</p> <p>○ <b>신 건설생산시스템 및 건설표준화기술</b> - 건설자재에 대한 이해향상 및 지정 모델화과정을 통해 표준화된 데이터베이스로부터, 설계와 자재선정, 구매 및 자재 재사용·재활용 등을 통해 건설 자재·부품의 활용범위를 확대하는 기술 (건축물 내·외부 마감공사의 자동화, 무인 자재운반 및 양중, 고층 건축물 외벽 유지관리 자동화 등 新 건설생산 시스템 기술 등 포함)</p> <p>○ <b>지능형 시설물 유지관리 기술</b> - 시설물 자동 위치정보의 활용으로 현장의 모든 장비에 대한 위치와 상태파악을 실시간으로 수행하도록 하는 첨단 유지관리 기술 (건설 시뮬레이션 및 모니터링 기술 개발을 통해 자동화 장비의 현장 또는 원격조정을 수행하고, 엔지니어링, 구매, 시공, 운영, 비상관리, 리노베이션, 철거단계 전반에 활용가능)</p> <p>○ <b>대심도 지하공간 구축·활용기술</b> - 도심 재개발 및 신도시 형성, 정보 인프라의 확충을 위해 대심도 지하공간을 구축·활용하는 기술 - 첨단 시공기술력 확보를 통한 공비·공기절감형의 신공간 창출 및 빛, 열, 공기, 소리, 위생 등의 환경요소와 화재 등 방재안전요소에 대한 공간 제어시스템 기술</p>	

## 중점 과학기술 번호 89

기술 분야	건설·교통·안전 분야
기술명	초장대교량 건설기술
기술 정의	<p>○ 초장대교량* 최적화 구조·설계, 대형기초 설계·시공, 고성능 재료적용, 유지관리 및 내풍성 향상기술 등이 집약된 고부가가치 건설 기술 (* 주탑 간의 거리인 경간장이 500m 이상으로 매우 긴 교량)</p>
세부기술 정의 및 범위	
<p>○ 초장대교량 설계기준 및 설계기법</p> <ul style="list-style-type: none"> <li>- 초장대교량 시스템의 구조 합리화 및 IT기술과 경량 신소재의 적절한 활용을 통해 경제성 확보와 함께 동적시스템의 안전성 제고가 가능하도록 하는 교량설계 관련 기술</li> </ul> <p>○ 내진/내풍에 의한 거동예측 및 진동저감기술</p> <ul style="list-style-type: none"> <li>- 풍동·지진 하중 등 동적하중에 대해 민감하게 반응하게 되는 초장대교량의 주형 강성 문제에 대해 구조물 자체의 형상 또는 부가적인 제어장치 등을 사용하여 구조물의 안정성을 확보하는 기술</li> </ul> <p>○ 첨단 시공기술 및 특수 장비 개발기술</p> <ul style="list-style-type: none"> <li>- 초장대교량 시공상의 여러 난제들을 해결하기 위한 첨단 시공·장비 관련 기술</li> <li>- 300~400m 높이의 사장교, 현수교용 고주탑의 설계·정밀시공 기술, 100~200m 이상 수심의 해상기초 설계·시공 및 심해 기초지반 개량설계·시공기술, 해상구간에서의 상부구조가설 및 케이블 가설을 위한 장비 및 공법(고속 Heavy Lifting 장비, Self mooring 해상 Barge, Multi- Strand Cable 등) 관련 기술 등이 포함</li> </ul> <p>○ 초고성능 재료 개발 기술</p> <ul style="list-style-type: none"> <li>- 초장대교량의 구조성능 확보를 위해 필요한 고성능 강재 이용기술 및 경량화·고강도 신소재 개발 등 재료개발 및 응용기술 (사장교·현수교용 고강도 케이블, 교량용 고성능 강재·콘크리트 구조재료 등 관련)</li> </ul>	

## 중점과학기술 번호 90

기술 분야	건설·교통·안전 분야
기술명	미래 첨단도시 건설기술
기술 정의	<p>○ 시민들의 웰빙(Well-Being) 수요의 증대에 따라 자연과 인간이 공생하는 새로운 개념의 생태도시에 대한 요구에 부응하고, 미래의 유비쿼터스 환경에 적합한 도심 속의 쾌적한 주거공간 건설을 위한 미래지향적 도시 건설기술</p>
세부기술 정의 및 범위	
<p>○ U-Eco City (친환경 도시건설 기술)</p> <p>- 21세기 미래형 첨단도시인 U-City 관련 요소기술 및 친환경 생태단지 구축을 위한 요소기술 개발을 통해 국민 주거생활의 편의성 및 쾌적성을 증진시키는 기술</p> <p>○ 복합공간 개발 기술</p> <p>- 지하 대공간과 인공섬 조성, 해저터널, 초고층 복합빌딩 기술 등의 기반기술 및 시범 추진을 통한 노하우 축적 등 미래 복합공간 개발을 위한 선도적 개발 기술</p> <p>○ 도시 재생 시스템 기술</p> <p>- 도시 건축물의 재생시스템 구축을 위하여 도시공간 구조의 재편, 도시 구조물의 순환 재활용, 신공간 창출을 도모할 수 있는 도시공간의 효율적 활용 및 공간 재창출 기술</p>	

## 중점과학기술 번호 91

기술 분야	건설·교통·안전 분야
기술명	지능형 국토지리정보구축기술
기술 정의	<p>○ 국토공간정보의 효율적 구축·가공·활용의 필요성이 높아짐에 따라, 기존 지리정보 구축에 중심에서 첨단 IT기술 및 GIS 관련기술을 활용한 사용자 맞춤형 공간정보 및 다양한 지능형 서비스를 제공하는 미래지향적 국토·도시공간 활용 기반기술</p>
세부기술 정의 및 범위	
<p>○ 국토 모니터링 기술</p> <ul style="list-style-type: none"> <li>- 국토 공간정보기술 활용을 통해 홍수해, 산불, 지진 등 국토에 발생하는 재해·재난을 실시간 감시·관리하여, 국민이 안전한 생활을 영위토록 하는 국토·환경·자원 모니터링 기술</li> </ul> <p>○ 공간정보 기반 인프라 기술</p> <ul style="list-style-type: none"> <li>- 향후 다양한 센서의 도입 및 공간정보 및 영상자료 정확도의 향상, 통합시스템의 도입 등을 통해 공간정보자료 취득, 국가안보 지도제작 등 다양한 분야의 자료축적 및 인프라를 구축하는 기술</li> </ul> <p>○ U-GIS 기반 건설정보화 기술</p> <ul style="list-style-type: none"> <li>- 유비쿼터스 국가정보화의 핵심인 국토 지리정보시스템(GIS) 관련기술을 SOC 건설과정에 유기적으로 통합 활용하여 건설기술을 고도화시키는 융합형 기술 (환경영향평가 및 교통영향평가 지원시스템 개발, GPS 이용 정밀시공 기술, SOC 유지관리기술 고도화, GIS 활용 수자원 관측기술 고도화 등)</li> </ul>	

## 중점과학기술 번호 92

기술 분야	건설·교통·안전 분야
기술명	미래 첨단 주거·교육환경 기술
기술 정의	<p>○ IT기술의 발전을 바탕으로 미래 생활환경에 있어서 주거공간의 환경 개선, 생산성 향상, 이용의 편리성 등을 추구하는 첨단 인텔리전트 빌딩, 홈 오토메이션 구축, 지능형 도서관 등의 정보통신 기술에 기반한 미래형 주거·교육 인프라 구축 관련 기술</p>
세부기술 정의 및 범위	
<p>○ 인텔리전트 교육시설(강의실, 도서관 등) 건설기술</p> <ul style="list-style-type: none"> <li>- 시간적·공간적 제약을 뛰어 넘어 언제 어디서나 국민에게 양질의 학습 시스템 제공을 위해 첨단 인텔리전트 교육시설(H/W) 구축 및 효율적 시설 운영·관리체계(S/W) 기술 (첨단 교육시설 Pilot Project 계획 및 구조설계 시스템, 원격학습 설비, 장애인 편의 시설 등 인텔리전트 설비구축, 고성능·다기능 교육환경 운영·관리 기술 등)</li> </ul> <p>○ 친실버 및 친환경 주거환경 구현 기술</p> <ul style="list-style-type: none"> <li>- 장애인, 노약자 등 사회적 약자의 자립생활 지원 및 사회통합에 기여할 수 있는 친실버·친환경적 생활환경 구현기술 (사회적 약자의 활동에 장애물이 없는 생활환경 설계·구현, 사회적 약자를 위한 주거 모델 개발기술, 환경 보존을 고려한 주민 교통편의 분석 기술 등)</li> </ul>	

## 중점과학기술 번호 93

기술 분야	건설·교통·안전 분야
기술명	미래 첨단 교통시스템기술
기술 정의	<p>○ 유비쿼터스 환경에서의 교통 환경 및 교통체계 효율화를 통해 국가 경쟁력 향상 및 국민 삶의 질 제고에 기여하는 교통 시스템 구현 기술. 교통혼잡 및 교통사고 비용의 획기적 절감 및 친환경적이고 효율적인 첨단교통시스템 구축을 위한 시스템 기술</p>
세부기술 정의 및 범위	
<p>○ ITS 기술</p> <ul style="list-style-type: none"> <li>- 21C 유비쿼터스 환경 하에 첨단 정보통신기술을 이용한 '교통체계의 효율화 및 첨단 교통운영관리 시스템 구축'을 위한 지능형교통시스템(Intelligent Transport Systems) 구축 및 표준화 기술</li> </ul> <p>○ U-Transportation 기술</p> <ul style="list-style-type: none"> <li>- 유비쿼터스 시대에 급격히 변화하는 교통환경에 대응하고 국민의 쾌적·신속·안전·편리한 차세대 대중교통 서비스를 제공하기 위한 新 교통시스템 개발 기술</li> </ul> <p>○ 미래 교통안전향상 기술</p> <ul style="list-style-type: none"> <li>- 교통 관련사고 및 재난을 통한 교통관련 사망자 및 부상자를 최소화 하고, 교통혼잡 비용을 절감시켜 효율적이고 환경친화적 국가교통시스템 구축 기술 (교통정보에 기반한 교통시설물 관리 및 교통관제기술, 차량 안전도평가 및 차량안전 기술, 자동단속기술, 교통안전시설 구축기술 등 포함)</li> </ul> <p>○ 교통연계 및 환승시스템 기술</p> <ul style="list-style-type: none"> <li>- 교통연계 및 물류의 체계화를 통해 고효율의 복합수송체계를 실현하기 위한 기술</li> <li>- 대중교통 이용 활성화를 위한 복합연계 및 환승 교통체계 구축 및 광역 대중교통체계를 전제로 한 도시설계 및 도로계획, 지하철 계획 등 전국적 교통연계 체제 구축 기술</li> </ul>	

## 중점 과학기술 번호 94

기술 분야	건설·교통·안전 분야
기술명	해양·항공운항 효율화 및 안전향상기술
기술 정의	<p>○ 글로벌 인적·물적 교류를 담당하는 항공 및 해양운항 분야의 범국가적인 운항 안전성 제고 및 효율화 시스템 구축 관련 기술.</p>
세부기술 정의 및 범위	
<p>○ 중소형항공기 인증기술</p> <ul style="list-style-type: none"> <li>- 미래의 항공운항 산업 개념에 큰 변화를 가져올 것으로 예측되는 소형제트 항공기의 등장과 함께, 이에 따른 차세대 입체교통망 기술 실현에 필수적인 중소형항공기 인증 기술.</li> </ul> <p>○ 항공 교통·운항 효율화 기술</p> <ul style="list-style-type: none"> <li>- 지속적으로 증가하는 국내 항공교통량을 원활하게 수용하고 보다 안전하고 경제적인 항공 운항여건 마련과 21세기 동북아 항공물류 허브로의 도약 등 국내외의 요구에 부응하기 위한 차세대 항행시스템(CNS/ATM) 및 운용지원 관련 기술</li> </ul> <p>○ 항공안전기술 / 해양안전기술</p> <ul style="list-style-type: none"> <li>- 21C 동북아 항공중심국 및 항공선진국 진입을 목표로 세계 G-10권 수준으로 향상시키기 위한 핵심 항공안전·인증기술의 개발·확보 및 범국가적 항공종합안전 시스템을 구축을 위한 지원기술</li> </ul> <p>○ 해양안전기술</p> <ul style="list-style-type: none"> <li>- 예기치 않은 선박, 항만, 해양구조물 등 해양 시스템 관련 사고를 통해 발생하는 인명, 재산, 환경 피해를 최소화하기 위한 기술 (선박안전, 운항안전, 해양관련 인적안전, 구난방재 및 해양재난 종합관리체계 구축 기술 등 포함)</li> </ul> <p>○ 해양교통관리기술</p> <ul style="list-style-type: none"> <li>- 날로 혼잡해지는 해양교통에서 대규모 해상재난 발생을 억제하고, 선박 사고에 의한 인명사고를 최소화시키기 위한 해양교통 관리 및 안전향상 기술 (해상 교통환경 정비 기술, 선박 안전운항 확보기술, 소형선박 안전성 확보기술, 해난 구조 체계 확립 기술 등 포함)</li> </ul>	

## 중점 과학기술 번호 95

기술 분야	건설·교통·안전 분야
기술명	차세대 고속열차 기술
기술 정의	<p>○ 남북통일과 동북아 시대에 대비한 안전성과 친환경성을 갖춘 첨단 미래형 고속열차의 속도 향상 및 무인화·경량화 관련 기술</p>
세부기술 정의 및 범위	
<p>○ 400km/h 분산형 고속철도기술</p> <ul style="list-style-type: none"> <li>- Super KTX 라 명명된(VC-10, 건설교통부, '06) 최고 운행속도 400km/h급 분산형 고속열차 시제차량 개발을 위한 핵심기술</li> <li>- 한국형 차세대 고속열차 실용화 시제차량 개발 및 실용화를 위한 시운전·시험·평가 등을 포함하는 고속철도 시스템 상용화 기술</li> </ul> <p>○ 초고속자기부상열차기술</p> <ul style="list-style-type: none"> <li>- 초전도자석을 이용하여 500km/h급으로 열차속도를 향상시키는 미래형 고속열차 속도향상 기술</li> </ul> <p>(초고속 자기부상철도 개발을 위한 핵심요소(초전도 코일, 전력공급장치 外)기술, Maglev 철도 시스템 기술 및 건설기술, 시험선 건설 및 운영관리기술 등 포함)</p>	



## 중점과학기술 번호 96

기술 분야	건설·교통·안전 분야
기술명	첨단도시형 철도시스템 기술
기술 정의	<p>○ 대중교통서비스의 질 향상과 국가 경쟁력 강화를 위한 새로운 도시철도시스템 개발기술 (기존 중전철 시스템에 신뢰성·안전성·운영효율성을 높인 차세대 전동차 시스템 기술, 자기부상열차, 경량전철 등 신 교통시스템 실용화 기술, 차량 전기신호선로 역사통신 등 도시철도시스템 표준화 기술 등을 포함)</p>
세부기술 정의 및 범위	<p>○ <b>첨단경량전철 및 신 교통시스템 기술</b></p> <ul style="list-style-type: none"> <li>- 지하철 차량보다 작은 규모로서 일정한 궤도를 따라 완전 자동 무인주행 구현이 가능한 신 철도교통 시스템 기술. (대도시에서 기존 지하철과 연계시켜 1~2분 이내의 짧은 배차시간 및 급기율·가·급커브 주행성을 높이고, 가속·감속 능력 향상을 통한 많은 도시민에 경전철 서비스를 제공하는 기술)</li> </ul> <p>○ <b>도시형 자기부상 시스템 기술</b></p> <ul style="list-style-type: none"> <li>- 저소음, 저진동의 환경친화적인 특성을 겸비한 도시형 자기부상철도 시스템(110km/h 급) 개발을 통해 대도시 수송시스템에 있어서 수송 분담률을 높인 보다 효율적 미래형 교통시스템 구축 기술 (전자기 흡입력 관련 기술, linear motor 추진시스템 등의 세부 기술 포함)</li> </ul>

## 중점과학기술 번호 97

기술 분야	건설·교통·안전 분야
기술명	첨단물류기술
기술 정의	<p>○ 무역 자유화와 기업활동의 글로벌화에 따른 물동량 증가 및 동북아시아의 급부상 등 급격히 변화하는 물류환경에 대응하고 물류강국으로 도약하기 위한 효율적 국가물류시스템 구축기술</p>
세부기술 정의 및 범위	
<p>○ <b>통합 지능형 컨테이너 시스템 기술</b></p> <ul style="list-style-type: none"> <li>- 해상·육상 등 다양한 운송수단을 이용하여 이동되는 컨테이너 화물의 안전한 수송과 선박, 장치장 등에 집적된 컨테이너의 센싱, 인식, 통신, 관제 및 운용기술이 통합적으로 결합·관리되는 시스템 기술 ( 컨테이너 내부 화물상태의 실시간 모니터링을 통한 사전 위험 감지·방재 기술 등 포함)</li> </ul> <p>○ <b>u-통관 시스템기술 (유비쿼터스 통합 물류정보화 기술)</b></p> <ul style="list-style-type: none"> <li>- Ubiquitous, Telematics, SMART CARD 등의 IT 신기술을 물류산업에 체계적이고 효과적으로 도입하기 위한 핵심기술 개발 및 표준화 기술</li> <li>- 기존 운영 중인 물류정보시스템간의 원활한 통합을 위한 연동기술 개발 등을 포괄하는 시스템 기술</li> </ul> <p>○ <b>지능형 교통물류 정보화 기술 (물류인프라 구축 및 운영 효율화 기술)</b></p> <ul style="list-style-type: none"> <li>- 물류 인프라 운영정보의 신속하고 효율적인 지원을 위한 시스템 기술 (지능형 물류 및 대량고속 화물 수송체계 기반 구축을 위한 상호연계형 통합물류, 복합운송지원형 연계운송 인프라, 고속대용량환경 친화적 물류 인프라 구축 기술 등 포함)</li> </ul> <p>○ <b>지능형 물류센터 및 고효율장비 기술 (Multi-modal Handling System)</b></p> <ul style="list-style-type: none"> <li>- 수송수단간 또는 물류거점내 원활한 연계를 통하여 복합·연계 물류체계의 효율을 극대화하기 위한 표준 운송용기 및 무인 화물처리를 위한 자동화 시스템의 설계·개발·적용기술</li> </ul> <p>○ <b>미래형 수송시스템 기술</b></p> <ul style="list-style-type: none"> <li>- 새로이 등장하는 물류운송·이동 시스템을 기반으로 글로벌화 된 물류 이동의 효과적 지원을 위한 통합된 물류수송 시스템 구축 기술</li> <li>- 산업 발전의 가속화·고도화 추세에 따라 물류수송이 이루어지는 곳의 지리적 특성 및 국토여건에 맞는 육상, 해상, 공중 등을 이용한 고속의 운송 시스템 구축 기술</li> </ul> <p>○ <b>해운물류기술</b></p> <ul style="list-style-type: none"> <li>- 첨단항만 건설 및 해운물류 처리시스템의 선진화를 위한 관련 기술, 해양공간의 개발·이용에 필수적인 각종 선박, 해양구조물 및 해양장비의 설계·해석·제작·운송·설치 관련 엔지니어링 기술, 해양 안전사고 예방 및 국민 안전을 위한 해양운송 관리·제어 시스템 기술 등을 포괄하는 기술</li> </ul>	