

우수 아파트단지 취재기사에서의 관리상의 논점 - 탐방기사를 이용한 언어통계학적 내용분석 -

Issues on Articles Covering Outstanding Management of Apartment Complexes - Content Analysis of Newspaper Reports with Lexical Statistics -

최정민* 강순주**
Choi, Jung-Min Kang, Soon-Joo

Abstract

Nowadays, diverse mass media discovers and introduces outstanding management cases of apartment complexes to induce vital competitions of constructors and active participation of residents to apartment management. This study statistically analyzed the management issues of outstanding apartment complexes that have been introduced by mass media with lexical criteria to examine the characteristics of their exemplary management. The key issues of outstanding apartment management are summarized as: efficient management of convenient facilities for residents, community activities based on residents' participation, and maintenance of pleasant living environments through transparent management. Also, the result of the relation arrangement of co-occurrence word from a Social Network Analysis included three key concepts of multi-family housing management - Maintenance Management, Operating Management, and Community Life Management - with emphasis on 'residents' and 'apartment complexes.' However, Operating Management was relatively deemphasized.

Keywords : Apartment Management, Newspaper Article, Content Analysis, Lexical Statistics, Text Mining

주요어 : 아파트관리, 신문기사, 내용분석, 언어통계학, 텍스트마이닝

1. 서론

1. 연구배경 및 목적

공동주택이 양적 공급에서 질적 관리의 시대로 접어들어 따라 집합주택 단지에서의 관리의 중요성이 점차 강조되고 있다. 이러한 움직임의 일환으로 1990년대 후반부터 지자체별로 우수관리단지 선정사업이 활발히 추진되고 있다. 우수 아파트관리 단지 사례를 발굴하여 널리 홍보함으로써 단기간 선의의 경쟁을 유도하고 입주민의 적극적 관리참여를 꾀함으로써 궁극적으로는 양질의 주택제고와 공동체 활성화를 도모함에 있다고 하겠다. 예컨대, '아파트 한가족 운동상'(부산시, 1995년), '아파트 우수관리단지'(경기도, 1997년), '살기 좋은 아파트'(인천시, 충북, 2000년), '아파트 관리 우수단지'(서울시, 2001년) 등이 그것이다. 한편, 살기 좋은 아파트 건설을 유도하기 위하여 1996년부터 '살기 좋은 아파트 선발대회'가 매년 실시·발표되고 있다.

이에 따라 매스컴에서 우수 아파트단지를 발굴하여 소개하는 경우가 최근 눈에 띄게 늘었다. 이러한 모범사례

는 소비자의 입장에서 본다면 아파트관리에 대한 관심과 참여를 유도하는 계기를 제공하며, 기사화된 해당 건설업체의 입장에서는 이를 마케팅 전략의 일환으로 적극 활용할 수 있는 기회를 제공한다. 마찬가지로 해당 아파트의 관리주체의 입장에서도 홍보효과에 따라 자치관리라면 집값 상승을, 위탁관리라면 수주전략 등을 기대할 수 있기 때문에 우수 아파트사례는 주목의 대상일 수밖에 없다.

본 연구는 이러한 배경을 바탕으로 매스컴의 우수 아파트단지 사례에서 소개하고 있는 모범적인 아파트 관리상에서의 논점을 언어통계학적 관점에서 계량적으로 분석하여 특성과 이에 따른 시사점을 고찰하고자 한다. 구체적으로는, 아파트관리 전문분야신문의 기사내용을 분석하는 일종의 내용분석인데, 방법에서 일반적 내용분석과는 다른 접근방법을 취하고 있다. 즉, '언어통계학적' 단어가 의미하듯, 기사내용에서 출현하는 단어의 출현빈도와 문장에서 특정 단어와 단어 간의 공출현 관계에 대한 언어의 통계량을 계산하여 문서(취재기사)의 논점 및 특성을 살펴보는 방법이다. 그동안 신문기사를 이용한 내용분석은 많은 학문분야에서 다양한 형태로 이루어져 왔

*정회원(주저자), 건국대 건축대학 주거환경 전공 조교수, Ph.D

**정회원, 건국대 건축대학 주거환경 전공 교수, Ph.D

1) 주최 : 매일경제 신문사, 후원 : 건설교통부, 역대 수상아파트 : <http://www.pugio.com>

나, 본고에서 접근하는 형태의 기법은 특정분야(언어학, 정보학, 사회학 등)에서만 선택적으로 응용되고 있을 뿐, 아직 주거 및 도시, 건축 등 건설 분야에서 시도된 사례는 찾기 힘들다. 그런 의미에서 본고는 기존의 주거관련 연구에서 사용된 연구방법과는 색다른 형태의 연구방법을 사용함으로써 새로운 연구방법의 모색이라는 관점에 분석체계와 기법내용에 대한 소개도 주요 연구내용으로 하였다.

2. 연구방법 및 분석체계

본 연구의 언어통계학(linguistic statistics, lexical statistics)적 분석방법과 관련된 학문분야로는 계산언어학, 텍스트 마이닝, 자연언어처리, 사회언어학 등 다양한데, 구체적인 분석방법에 대해서는 제2장에서 자세히 언급한다. 본 연구에서의 분석체계와 흐름은 <그림 1>과 같다.

본 연구의 조사도구로 사용된 신문기사는 (주)아파트관리신문사²⁾에서 온라인 공개하고 있는 우수 아파트관리 단지의 탐방 기사를 이용하였다. 이 신문은 아파트관리에 관한 전문분야신문으로서 역사가 길고, 또한 탐방기사는 장기간 고정적으로 모범사례를 소개하고 있으므로 본 연구의 분석에 적합하다고 판단하였다. 탐방기사에 대하여 간략히 소개하면, 원칙적으로 매주 한편씩 발행되고 있는데, 1994년 12월 24일 제1호부터 2006년 3월 27일 현재 623호까지 발행되어 있다. 5명의 전문기자가 각 지자체 등에서 실시하고 있는 아파트관리우수단지(서울시의 경우) 또는 아파트관리 분야 모범사례를 발굴하여 취재·소개하고 있다. 취재 대상단지는 원칙적으로 아파트 의무관리

대상인 최소 150세대 이상, 주택관리사(보)의 배치가 의무화되어 있는 곳이기 때문에 일정규모 이상의 민간 분양아파트단지가 대부분이다.

그런데 본 연구의 분석방법을 적용하기 위해서는 전자 텍스트 기사이어야 하는데, 기발행 탐방기사 중에서 현재 텍스트 추출이 가능한 것은 329호(1999.12.25)~619호(2006.2.27)만이였다. 그 결과 내용이 불비한 9개의 기사를 제외시켜, 최종적으로 282개의 기사를 수집하여 분석에 이용하였다. 전자 텍스트 기사를 대상으로 하므로 문서중의 사진(이미지)과 기사이름, 날짜 등 연구분석에 불필요한 내용을 제외하고 하나의 텍스트 파일형태로 만들어 집계하여 보았더니 그 규모는 개략 1.08 MB의 크기였다³⁾.

한편, 추출한 282개의 본문 기사는 대용량의 문서자료(텍스트마이닝의 전문용어로는 코퍼스(corpus)⁴⁾라고 함)로서 통상의 내용분석을 통하여 논점을 추출하는 작업은 적잖은 비용(cost)을 필요로 하기 때문에 반자동적인 방법을 모색하였다. 그 결과 2장에서 소개하는 텍스트마이닝의 접근방법을 시도하게 되었고, 이 과정에서 주거학을 포함하여 건설분야에서 매스컴의 기사를 내용분석하고 있는 기존 연구를 리뷰하여 어떤 내용을, 어떤 관점에서 고찰하고 있는가를 살펴보았다.

다음으로, 텍스트마이닝을 포함한 본고의 연구방법에 대하여 자세히 소개한 다음, 제시한 기법을 구체적으로 적용하여 언어통계 정보를 추출하고 그 함의를 분석하였다. 적용과정에서 전체의 본문기사(이하 'FT기사'), 큰제목인 헤드라인 기사(이하 'HD기사'), 그리고 중간제목 기사(이하 'ST기사')의 세 가지 형태로 나누어 추출하여, 유형별 논점을 세부적으로 분석하였다.

II. 언어통계학적 분석과 텍스트마이닝

1. 내용분석과 언어통계학적 분석방법

1) 내용분석과 언어통계학적 분석

신문기사 등을 이용한 전통적인 내용분석(content analysis)⁵⁾에서는 분석대상이 되는 자료를 적정량 수집하여 분석목적에 따라 수량적 분석 또는 질적 분석을 수행한다. 그런데 전통적인 내용분석 기법이 많은 기여를 했음에도 불구하고 다음과 같은 한계가 지적되고 있다. 첫째, 연구자가 임의로 만든 분석항목(category)에 너무 의존적이라 객관성이 결여될 수 있고 둘째, 외적 타당성

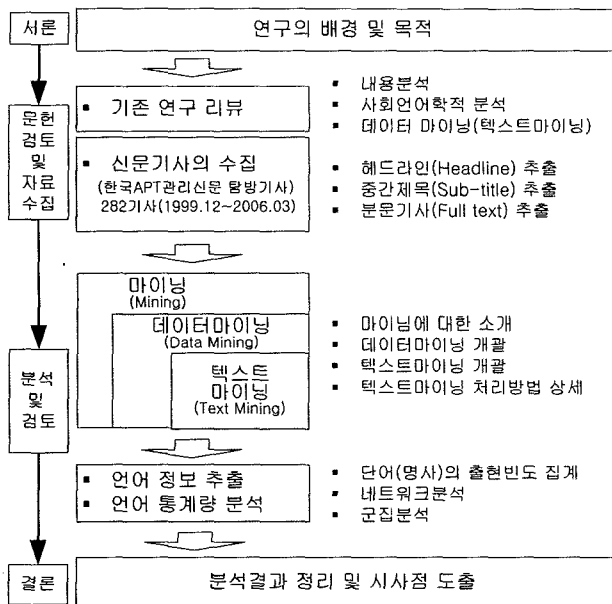


그림 1. 연구의 분석체계와 흐름

2) 1993년 2월 창간: <http://www.aptn.co.kr/>. 본 연구 분석을 위하여 저작권자로부터 학술적 이용의 허가를 득하였다.

3) MS Word에서 계산했을 경우, 단어수로는 약 143,600단어, 문자수로는 약 485,680자(공백제외)에 해당한다.

4) 언어를 연구하는 각 분야에서 필요로 하는 연구 자료로서, 언어의 본질적 모습을 총체적으로 드러내 보여 줄 수 있는 자료의 집합을 뜻한다. 한국어로는 말뭉치 또는 말모듬으로 번역하는데 그 정의는 사람에 따라 다르다. http://transkj.com/corpus_is.htm

5) 메시지로써 표현되는 특정 심벌, 명제, 인물 등을 분석단위로 하고 그들에 대한 출현빈도의 계산, 출현공간의 계측 및 평가 등 미리 설정한 카테고리화 판단기준에 따라 분석한다.(네이버:<http://100.naver.com/>)

이 제한되어 있어 연구자의 정치적(가치관, 선입감 등) 성향에 영향을 받을 수 있고 셋째, 대량의 자료를 분석하기가 용이하지 않아 비용이 많이 든다는 점 등이다(박한우, 2004; 崔廷敏, 2004). 특히, 분석과정에서 연구자의 자의적 판단에 따라 전형적이라고 생각되는 사례만을 인용하거나 해석하는 경우가 있다. 그런데 문제는 해당 사례가 전형적이라고 말할 수 있는 근거, 기준이 애매한 경우가 있는데, 이에 대하여 Mida(見田, 1965)는 유형에 관한 범위, 차원 또는 요인의 범위, 요인의 현시성이라는 세 요소를 지적하고 있다. 이러한 전통적 방법이 지닌 문제점을 극복하기 위해서 근래 들어 컴퓨터 및 정보처리 기술의 발전에 힘입어 계산기 의존적인 방법이 다양하게 시도되고 있다. 본 연구에서도 이러한 배경에 따라 자연언어처리 기술을 주로 응용하고 있는 텍스트마이닝의 접근에, 언어통계학적 분석기법을 연계함으로써 주거학 관련분야에 실험적으로 시도하려는 것이다.

2) 내용분석 및 마이닝 관련연구의 리뷰

위에서 언급한 전통적 내용분석 기법을 이용하여 건설 관련 분야에서 적용한 사례를 찾아보면, 신문과 인쇄매체의 광고를 이용한 사례⁶⁾, 잡지나 대중소설을 이용한 연구 사례⁷⁾, 일간지의 신문기사를 이용한 사례⁸⁾ 등 다양하다.

한편, 대조적으로 본 연구에서 접근하고 있는 언어통계학적, 또는 텍스트마이닝의 기법을 이용한 건설분야에서의 국내 적용사례는 거의 보이지 않는다. 그 이유는 건설관련 전공자가 일련의 자연언어처리 과정을 프로그래밍하기가 용이하지 않은데다가 언어학적 이해가 부분적으로 필요하고, 무엇보다도 건설관련 분야에서 적용된 사례를 찾기가 용이하지 않아 적용가능성이 가시화되지 않았기 때문인 것으로 판단된다. 그러나 텍스트마이닝을 부분적으로 포괄하는 형태라고도 할 데이터마이닝의 기법을 이용해서 건설관련 분야에 적용한 사례는 최근 꾸준히 증가하고 있는 추세이다. 예를 들면, 도시계획 분야에서는 임차인의 임대차 계약 선택과 도시성장 모델링에 적용한 사례⁹⁾, 건축 분야에서는 시공관리에 적용된 사례¹⁰⁾ 등이 부분적으로 보이고 있다. 그런데 전통적으로 데이터마이닝의 응용사례가 풍부하고 또한 이 응용기술이 발전하여온 도양은 고객관계관리(CRM)와 연계한 마케팅 분야라고 할 수 있을 것인데, 아직 국내 건설관련 마케팅 분야에서의 적용 사례는 학술적으로 보고되고 있지 않다.

2. 데이터마이닝과 텍스트마이닝

1) 마이닝

마이닝(mining)이란 무엇인가를 발견하는(찾아내는) 것

을 의미하는 것으로 학문적으로는 일반적으로 ‘발견과학’으로 지칭되고 있다. 이 단어가 의미하는 것에서 알 수 있듯이 그 대상 소재인 데이터는 방대하여야 하고 추출한 事象(규칙이나 패턴 등)은 발견에 걸맞아야 하는 등 몇 가지 파생적인 성질(예컨대, 신규성이나 유용성)을 가져야 한다(崔廷敏, 2004). 이러한 ‘발견’이라는 知的 행위는 언제나 사람들을 매료시켜 왔으며 모든 분야에서 인류 진보의 원동력이 되어 왔다. ‘발견’과 ‘학습’이 가지는 의미는 다르나 역사적으로 발견은 관측으로부터 학습이 어려운 형태라고 간주되어 왔다(Langley, 1986).

마이닝 관련연구는 최근 정보기술의 눈부신 발전에 편승하여 학제적(interdisciplinary)¹¹⁾으로 이루어지고 있으며 응용기술과 적용분야도 폭넓다. 그런 까닭에 특성상 현재 진행형의 연구 분야라 할 수 있으며, 따라서 연구체계가 확립되어 있다기보다는 구축되어 가고 있다는 것이 올바른 표현일지 모른다.

2) 데이터마이닝(Data Mining)

데이터마이닝이란 통상 대량의 데이터로부터 규칙성이나 관련성 등 의미 있는 패턴을 자동적으로 추출하는 수법이라고 정의된다(Adriaans et al., 1998; Bounsaythip et al., 2001; Choi et al., 2002). 유사한 용어로서 KDD(Knowledge Discovery in Databases; 지식발견), 정보추출(Information Extraction)등이 있다. 주택 또는 도시·건축 분야에서도 ‘정보’를 테마로 한 연구와 데이터는 최근 급증하고 있으나 앞서 설명한 바와 같이 데이터마이닝 기술을 이용한 연구는 현재 그렇게 알려져 있지 않다. 이 중에서 주택분야에서 이루어진 연구로는 崔·淺晃(2003)가 있는데, 여기서는 일본 주택시장에서 대량의 통계 데이터로부터 흥미 있는 정보만을 추출하는 새로운 수법을 제안하고, 이를 실제 신규분야아파트 계약자 데이터베이스에 적용하면서 그 유용성에 대해서 논하고 있다. 주택분야에서의 마이닝 기술은 마케팅 분야와 접목함으로써 다양한 응용예가 소개되고 있는데, 앞서 CRM 등을 중심으로 방대한 고객 데이터로부터 고객 특성정보를 추출하여 이를 마케팅 활동과 연계하는 형태로 응용되고 있다¹²⁾.

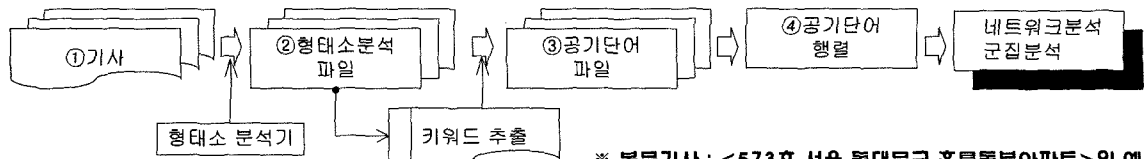
한편, 대상이 되는 데이터의 형식도 통계데이터, 문서데이터, 웹데이터, 인터넷의 입출력데이터, 공간(spatial)데이터 등 다양하다. 말하자면, 이용하는 데이터 형식에 마이닝 기술을 접목하여, 예컨대 웹마이닝(인터넷 웹데이터 이용), 텍스트마이닝(문서데이터 이용), 공간데이터마이닝(공간데이터 이용) 등으로 부르고 있다.

3) 텍스트마이닝(Text Mining)

앞서 설명한 바와 같이 텍스트마이닝은 문서데이터(텍스트 파일)를 이용하여 유용한 정보를 마이닝(추출)하는

6) 이재혁·제해성(2005), 홍윤영 외 2인(2004), 손세관·김승언(2000), 신용재(1996), 신창현(2003) 등
 7) 박철수(2005), 姜滙 외 4명(1999) 등
 8) 이성미·이연숙(2004), 高橋由光 (2002) 등
 9) 김태운·이창무(2005), 김정엽 외 2인(2004) 등
 10) 이종국(2004), 박우열(2005) 등

11) 주요 관련학문 분야로서는 데이터베이스, 기계학습, 통계학, 패턴인식 등을 들 수 있다.
 12) 예를 들면, Bounsaythip & Rinta-Runsala(2001), 松田(2001)



※ 본문기사 : <573호 서울 동대문구 홍릉동부아파트>의 예

①본문기사	②형태소분석 파일	③공기단어 파일	④공기단어 행렬																																																																				
<ul style="list-style-type: none"> 아파트는 수많은 이웃들이 모여 사는 마을로 공동의 이익과 쾌적한 생활여건 조성을 위해 규칙·규정을 반드시 제정해야 하고 이를 철저히 준수할 수 있도록 유도해야 한다. 그러나 이를 관리규약에 모두 명시하기 힘들기 때문에 서울시 등 각 시·도는 사인별 관리규정을 통해 이를 구체화하도록 유도하고 있다. 지난해 서울 동대문구 우수단지 상을 수상한 바 있는 홍릉동부아파트(4개동 3백71세대, 쌍림건설산업(주) 관리)는 이적한 취지로 관리 외 수의 관리규정, 계약사무처리규정, 관리규정 등의 다양한 규정을 제정해 이를 준수하고 있다. 이는 단지 내 문화예술포럼, 관리업무, 원활한 공동생활에 큰 효과를 발휘하고 있다. ... 	<p>아파트는 (N "아파트"<N:20> + (j "는")<1> 수 많은 (V "수많은"<J:20> + (e "은") 이웃들이 (N "이웃"<N:20> + (s "들") + (j "이") 모여 (V "모이"<J:24> + (e "어")<8> (N "모")<NUM:31> + (j "여")<3> 사는 (V "사"<T:20> + (e "는") (V "살"<TJ:20> + (e "는") (V "시"<T:24> + (e "어")<4> (N "시"<NUM:27> + (j "는")<1> 마을로 (N "마을"<N:24> + (j "로")<2> 공동의 (N "공동"<N:20> + (j "의") 이익과 (N "이익"<N:20> + (j "과") 쾌적한 (V "쾌적하"<J:24> + (e "는")<13> (N "쾌적"<N:29> + (l "하") + (e "는")<13> 생활여건 (N "생활여건"< :53> 조성음 (N "조성"<N:20> + (j "음")</p>	<ul style="list-style-type: none"> 아파트, 이웃, 마을, 공동, 이익, 쾌적, 생활여건, 조성, 규칙, 규정, 제정, 준수, 유도 관리규약, 모두, 명시, 때문, 서울시, 사인별, 관리규정, 규범, 화, 유도 지난해, 서울, 동대문구, 우수단지, 수상, 홍릉동부아파트, 4개동, 3백71세대, 쌍림건설산업, 관리, 취지, 관리, 수의, 관리규정, 계약사무처리규정, 관리규정, 다양, 규정, 제정, 준수 단지, 분정예방, 효율적, 관리업무, 원활, 공동생활, 효과, 발휘 	<table border="1"> <tr> <td></td> <td>1.</td> <td>2.</td> <td>3...</td> </tr> <tr> <td>1. 입주민</td> <td>0</td> <td>0</td> <td></td> </tr> <tr> <td>2. 단지</td> <td>612</td> <td>0</td> <td></td> </tr> <tr> <td>3. 관리</td> <td>285</td> <td>188</td> <td></td> </tr> <tr> <td>4. 부녀회</td> <td>243</td> <td>270</td> <td></td> </tr> <tr> <td>5. 입주자대표회의</td> <td>319</td> <td>211</td> <td></td> </tr> <tr> <td>6. 관리직원</td> <td>220</td> <td>155</td> <td></td> </tr> <tr> <td>7. 관리주체</td> <td>279</td> <td>181</td> <td></td> </tr> <tr> <td>8. 공사</td> <td>169</td> <td>132</td> <td></td> </tr> <tr> <td>9. 관리비</td> <td>138</td> <td>100</td> <td></td> </tr> <tr> <td>10. 관리사무소</td> <td>166</td> <td>108</td> <td></td> </tr> <tr> <td>11. 세대</td> <td>143</td> <td>83</td> <td></td> </tr> <tr> <td>12. 대표회의</td> <td>108</td> <td>67</td> <td></td> </tr> <tr> <td>13. 시설물</td> <td>98</td> <td>119</td> <td></td> </tr> <tr> <td>14. 직원</td> <td>89</td> <td>68</td> <td></td> </tr> <tr> <td>15. 주거환경</td> <td>118</td> <td>99</td> <td></td> </tr> <tr> <td>...</td> <td></td> <td></td> <td></td> </tr> </table>		1.	2.	3...	1. 입주민	0	0		2. 단지	612	0		3. 관리	285	188		4. 부녀회	243	270		5. 입주자대표회의	319	211		6. 관리직원	220	155		7. 관리주체	279	181		8. 공사	169	132		9. 관리비	138	100		10. 관리사무소	166	108		11. 세대	143	83		12. 대표회의	108	67		13. 시설물	98	119		14. 직원	89	68		15. 주거환경	118	99		...			
	1.	2.	3...																																																																				
1. 입주민	0	0																																																																					
2. 단지	612	0																																																																					
3. 관리	285	188																																																																					
4. 부녀회	243	270																																																																					
5. 입주자대표회의	319	211																																																																					
6. 관리직원	220	155																																																																					
7. 관리주체	279	181																																																																					
8. 공사	169	132																																																																					
9. 관리비	138	100																																																																					
10. 관리사무소	166	108																																																																					
11. 세대	143	83																																																																					
12. 대표회의	108	67																																																																					
13. 시설물	98	119																																																																					
14. 직원	89	68																																																																					
15. 주거환경	118	99																																																																					
...																																																																							

그림 2. 공기단어 통계량에 의한 단어(키워드)의 네트워크분석 및 군집분석 처리과정

것이지만, 통상 관계형 데이터베이스에 저장되어 있는 정형의 데이터를 분석하는 것과는 달리, 특정의 데이터 구조를 갖지 않는 텍스트 문서를 대상으로 하고 있다는 점이 특이하다. 텍스트마이닝은 일반적으로 정보추출, 문서검색, 문서분류, 문서유형화 등의 요소기술을 조합하여 대량의 문서데이터로부터 새로운 지식을 획득하는 것을 지향하고 있다(市村·鈴木, 2001). 어쨌든 ‘언어’를 매개로 하는 응용연구 분야이지만 다방면의 학문분야와 관계되어 있는데, 전형적으로는 언어의 의미론, 심리론, 통계론, 사회론 등과 연계되어 있다. 그래서 ‘언어의 통계량’이라고 할지라도 언어학에서 다룬다면 언어통계학(lexical statistics)이나 계산언어학(computational linguistics)에 가깝고, 만약 자연언어처리(NLP; Natural Language Processing)라는 관점에서 다룬다면 정보공학(information engineering)에 가깝다. 또한 사회학의 관점에서는 사회언어학(sociolinguistics)과 언어사회학(sociology of language)¹³⁾이 있다. 지면관계상 텍스트마이닝에 관심 있는 독자는 Lebart et al.(1997)이나 Mannig et al.(1999) 등을 참조하기 바람이며, 다음에서는 본 연구에서 이용하는 언어통계 처리방법에 대하여 구체적으로 소개한다.

3. 본 연구에서 접근하는 텍스트마이닝 기법

1) 문서에서 명사의 추출

본 연구에서의 일련의 언어통계 처리방법을 <그림 2>

13) 사회언어학은 언어 중에서 사회계층이나 지역과 같은 언어 외적인 것이 어떻게 영향을 미치며 언어가 어떻게 사용되거나 변용되는지를 주로 분석하는 것이고, 언어사회학은 언어 자체의 발생론적 논의나 또는 모든 언어를 하나의 사회적 제도로서 생각하는 입장에 그 차이가 있다.

에 나타내었다. 우선, 본고에서는 문서에서의 언어 통계량의 단위로써 ‘단어(word)’에 주목하며, 단어 중에서 ‘명사’¹⁴⁾의 출현빈도¹⁵⁾를 이용하여 특성을 고찰한다. 한국어의 경우 “문장의 각 단어는 띄어 쓰임을 원칙으로 한다.”는 한글 맞춤법에 따라 표기되므로 문장에서 단어를 추출할 때는 공백(space)을 기준으로 단어를 추출하면 된다. 그런데 주어진 문장에서 명사만을 추출해서 출현빈도를 집계하는 일이란 생각만큼 그렇게 간단한 일이 아니다. 예를 들면, “주거학은 주거 생활공간과 환경을 거주자 측면에서 접근함과 동시에...”라는 문장에서, ‘주거학은’의 경우 ‘주거학’(명사)+‘은’(주격 조사)의 형태로 조합되어 표기되므로 명사인 ‘주거학’만을 추출하기 위해서는 분해를 해보아야 한다. 즉, 문장에서 문자열을 의미 있는 최소단위(이것을 통상 ‘형태소’라고 부르며, 형태소의 분석처리를 ‘형태소 해석’이라고 함)로 분해할 필요가 있다. 이렇게 문장을 형태소별로 분해하기 위해서는 통상 ‘형태소 분석기’라고 불리는 해석 시스템을 이용하여 수행한다. 이것에 대해서 간단히 말하면, 구문해석에 의해 얻어진 해석결과를 바탕으로 형태소별로 이미 등록되어 있는 단어 데이터베이스로부터 분해결과를 자동적으로 대응시킴으로써 문장에서 분절된 형태소를 특정하여 의미부여하

14) 품사에는 다양한 종류가 있지만 키워드로 제시되는 품사는 일반적으로 명사이기 때문에 여기서도 명사(보통명사, 집합명사, 추상명사, 물질명사, 고유명사)만을 대상으로 하였다.

15) 출현빈도를 통계량으로 이용하는 이유는, (1) 흥미를 가지고 있는 것, 또는 관심이 높은 것에 대해서는 해당 단어를 다루는 경향이 있다는 점, (2) 전달자는 몇 번이나 같은 단어를 반복함으로써 전달하고자 하는 내용을 강하게 의식시킨다는 점, 거꾸로 청취자의 입장에서는 같은 단어의 반복에 의해 기억에 강하게 남는다는 점 등을 들 수 있다.

는 것으로 이해하면 된다. 본고에서는 'KLT(ver 2.1.0b)'¹⁶⁾ 한국어 형태소 분석기(Korean morphological analyzer)를 사용하였다.(그림 2의 ①→②)

2) 공기(共起)단어의 집계

형태소 분석에서 얻은 결과파일(이하 '형태소분석 파일')에서 대상으로 하고 있는 명사만을 자동으로 추출하여 출현빈도를 집계하기 위해서는 별도의 프로그래밍 코드가 필요하다. 본고에서는 이를 위해 Perl¹⁷⁾ 프로그래밍 언어에 의한 자작 프로그래밍 코드를 이용하였다.(그림 2의 ②→③)

다음으로, 이 과정에서 추출한 명사를 출현빈도를 기준으로 하여 내림차순으로 정렬하고, 이 중에서 본문분석(FT기사)에서는 본 연구의 관심 대상인 '아파트관리'와 관계가 깊다고 판단되는 단어(명사)만을 상위 k까지 키워드로 추출한다. 키워드로는 어떤 단어가 적당한가 하는 문제는 연구 분야와 목적에 따라 달라지는데, 이는 달리 말해 단어에 대하여 어떠한 성격 매김을 행할 것인가의 문제에 귀착한다. 본고에서의 접근방법은 출현빈도를 우선시하면서 필자 등의 전문성을 감안하여 단어의 '중요성'과 '대표성'이라는 일반론적 관점에서 상위 k=100개의 단어를 추출하였다¹⁸⁾.

한편, 문서에서 관련이 깊은 단어끼리는 서로 가까운 위치에 빈번히 출현한다는 언어학적 '共起性(co-occurrence)'에 착안하여 단어 간의 밀접관계를 통하여 전달하려는 메시지내지 像(image)을 고찰하고자 한다. 공기단어 분석이란 문장(sentence)이나 단락(paragraph) 등 처리해야 할 단위문자열(전문용어로 'window'라고도 부름)에서 공통적으로 출현하는 단어의 빈도를 집계하여 그 언어학적 성질을 고찰하는 것이다. 본고에서는 이를 언어의 사회학적 특성과 연계함으로써 아파트관리 전문분야신문의 탐방 연재기사에서 소개되고 있는 우수 아파트단지에서의 관리상의 논점을 언어의 통계량에 기초하여 고찰하려는 것이다. 공기단어란, 쉽게 말하면 '주거'를 예로 하였을 경우 '주거문화', '주거계획', '주거환경', '주거복지' 등과 같은 단어들이 후보라고 할 수 있을 텐데, 왜냐하면 주거와 깊은 관련이 있어서 문장에서 '주거'와 함께 자주 출현할 단어이기 때문이다.

한편, 본고에서는 공기단어 분석의 window 단위로 '문장'을 설정하였는데, 이는 신문의 집필과정에서 취재자의 개별적 특성이나 습관 등을 가능한 배제하기 위

함이다. 왜냐하면 단락은 문장보다 길기 때문에 집필자의 은유(metaphor)와 습관 등이 개재될 여지가 많기 때문이다. 다음에는 이를 다변량분석(본고에서는 군집분석으로 한정) 또는 네트워크분석과 연계하기 위해서는 '공기단어 행렬'을 만들어야 한다. <그림 2>의 ④에서처럼 k개의 키워드로 이루어지는 대칭행렬(symmetric matrix)을 만들고, 대각행렬은 0(zero)으로 처리하였다. 이 과정은 R언어¹⁹⁾에 의한 프로그래밍 코드로 처리하였다(그림 2의 ③→④).

3) 네트워크분석 및 군집분석

상기 공기단어 행렬이 만들어지면 이 통계량에 기초하여 다양한 분석이 가능하다. 먼저 네트워크분석인데, 여기서는 사회학 등에서 많이 이용되고 있는 방법 중의 하나로 그래프이론(graph theory)에 기초한 '사회네트워크 분석'(Social Network Analysis)기법을 이용한다. 전통적으로 사회학에서 많이 이용되는 이 기법은, 어떤 사회적 시스템의 구조를 구성요소들이 맺고 있는 '관계'(relation)의 형태로 고찰하는데, 관계의 유형은 가시적인 것에서부터 비가시적인 것까지를 포괄한다. 이 방법을 언어통계학의 공기단어 분석에 유추(analogy)하면, 단어(키워드) 간의 의미론적 연관관계(semantic association)가 가시화되므로 문서에서의 논점을 중심어의 연결 관계로 재구성함으로써 연구의 관심인 '논점'을 정리할 수 있다. 왜냐하면 앞서 '주거' 단어의 예제에서 보인 바와 같이 문서에서 이 단어와 함께 문장단위로 공출현하는 '주거환경', '주거복지' 등과 같은 후보 공기단어간의 관계는 출현빈도의 정도로 구성할 수 있기 때문이다.

다음으로, 군집분석(cluster analysis)을 이용하여 공기단어간의 군집에 의한 유형화를 수행하여 군집된 유형별 특성을 살펴본다. 군집분석은 실로 다양한 기법이 제안되어 있지만 여기서는 사회과학에서 많이 사용되는 전통적인 위계적 군집분석²⁰⁾을 이용하고, 이 과정에서 공기단어 간의 출현빈도를 그대로 근접척도(proximity measure)로 이용하였는데, 이렇게 되면 출현빈도가 높은 단어일수록 군집위계(clustering order)가 높아져 덴드로그램(Dendrogram)상의 상부에 위치하게 된다.

4. 분석대상 기사의 특징과 구성

본 연구에서 분석대상으로 하고 있는 탐방기사에 대해서는 서론에서 간단히 설명하였으나, 분석의 이해를 돕기 위하여 이 자료에 대하여 설명을 추가할 필요가 있

16) 저작권자 및 출처: 강승식(국민대학교 언어 공학, 정보 검색 연구실), <http://nlp.kookmin.ac.kr/>

17) Perl(Practical Extraction and Report Language)은 인터넷의 급속한 보급에 의해 웹서버에서 외부 프로그램을 실행하기 위한 방법의 하나인 CGI(Common Gateway Interface)를 실행하기 위한 프로그래밍의 하나이다. 특히, 문자열의 패턴매칭(pattern matching)을 표현하는 '정규표현'의 처리에 뛰어나다.

18) HD기사 및 ST기사의 분석에서는 키워드를 별도로 추출하지 않고 출현빈도순으로 상위 k=100개의 단어만 추출하였다. 이들 기사에서는 불용어가 상대적으로 적기 때문이다.

19) R언어는 통계해석 언어 S의 클론(clone)으로 GNU의 규정에 개발이 추진된 통계해석 언어인데, 무료이면서도 고성능을 가져 웬만한 표준적인 통계수법은 거의 모두 간단한 명령어로 실행시킬 수 있다. 무엇보다도 공개 패키지가 풍부하며 부족한 기능은 패키지의 형태로 보완할 수 있을 뿐만 아니라 또한 자작 코드에 의하여 다양한 기능의 구현이 가능하다.

20) 여기서는 유클리드 '평균' 거리에 기초한 위계적 군집분석의 방식을 채택하였다(hierarchical clustering).

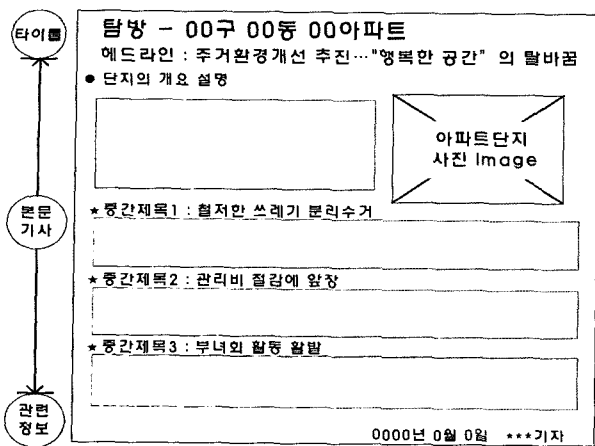


그림 3. 탐방기사의 표준 구성 레이아웃

다. <그림 3>은 본 연구에서 이용하고 있는 탐방기사의 전형적인 모습을 보여주고 있다.

우선 전체적인 구성은 “타이틀/본문 기사/관련정보”의 세 부분으로 나누어진다. 여기서 본문 기사는 다시 “헤드라인/단지개요+단지전경(이미지)/{중간제목1+관련기사1}, {중간제목2+관련기사2}, {중간제목3+관련기사3}”으로 분류된다.

한편, 분석대상으로 하고 있는 282개 취재기사의 지역별 분포를 집계해서 나타낸 것이 <표 1>인데, 표에서 알 수 있듯이 취재 지역은 대부분 서울과 경기, 인천의 수도권 지역이며, 지방은 10개 단지로 전체의 3.5%에 불과하다.

III. 언어통계학적 분석

1. 헤드라인(HD기사) 분석

1) 헤드라인 분석의 의미

언론정보학 측면에서 헤드라인(큰제목)은 수용자에게 전달 희망 내용을 핵심적으로 축약한 것으로 중요하게 처리된다. 이 과정에서 유사한 의미의 단어나 혹은 간략하게 표현할 단어는 전달자의 전문적 판단에 따라 결정된다(박한우, 2004). 이러한 헤드라인은 “① 독자의 관심을 끈다. ② 스토리를 요약한다. ③ 뉴스를 등급화 한다. ④ 신문을 매력적으로 만든다.”는 역할을 가지고 있다(이상우, 1982). 이에 따라 본 연구에서도 어떠한 내용들이 모범 아파트관리 사례에서의 헤드라인으로 소개되고 있는지를 살펴보았다. <표 2>는 본 연구의 샘플 헤드라인 기사를 예시한 것이다.

2) 출현빈도 분석

282개 헤드라인만의 추출문서에 대하여 앞에서 설명한 방법에 따라 명사만을 추출하여 출현빈도에 따라 내림차순으로 정리해보면 <표 3>과 같다. 출현빈도 상위 10단어를 보면, “입주민, 아파트, 관리, 조성, 주거환경, 쾌적, 실현, 투명, 시설물, 철저”의 순이다. 이 중에서 ‘명사+하다’²¹⁾의 형태로, ‘하다’를 붙여 명사의 의미 본질을 동

표 1. 연구대상 아파트단지의 지역별 분포(취재 단지 수)

지역	구/시/군	단지 수
서울	강남구	4
	강동구	8
	강북구	1
	강서구	9
	관악구	8
	광진구	1
	구로구	3
	금천구	2
	노원구	11
	도봉구	4
	동대문구	4
	동작구	10
	마포구	2
	서대문구	1
	서초구	1
	성동구	6
	성북구	6
	송파구	8
	양천구	3
영등포구	3	
용산구	2	
은평구	2	
중랑구	5	
경기	고양시	21
	과천시	2
	광명시	5
	광주시	2
	구리시	3
	군포시	6
	김포시	3
	남양주시	3
	부천시	11
	성남시	15
경북	수원시	19
	시흥시	11
	안산시	4
	안양시	11
	오산시	1
	용인시	5
	의왕시	3
	의정부시	5
	파주시	2
	평촌시	1
인천	평택시	4
	하남시	1
	경산시	1
	경주시	1
	대구 북구	1
	대전 서구	2
	부산 영도구	1
	울산 남구	1
	부산 북구	1
	울산 남구	1
전남	여수시	1
	군산시	1
전북	광주 남구	1
	광주 남구	1

표 2. 아파트관리신문의 샘플 헤드라인 기사

계재 번호	헤드라인 기사
564호	철저한 시설물 관리로 ‘장수명 아파트’ 만들기 앞장서
563호	입주민·관리직원 한마음으로 쾌적한 주거환경 만들어
562호	입주민간 교류 통해 정이 넘치는 아파트 만들어
548호	‘3C 운동’ 펼치며 쾌적하고 살기 좋은 아파트 조성
544호	농산물 직거래 등 실시 ... 공동체 문화 꽃피워
538호	입주민에 감동 주는 ‘감성관리’ 실현 ... 주거만족도 높여
534호	자연친화형 주거공간 조성해 입주민 만족도 높여
530호	이웃사랑 실천 ... 따뜻한 아파트 공동체 만들어
499호	상세한 관리계획 수립 ... ‘예방관리’ 실현 모범 보여
486호	이웃사랑 실천...ESCO사업 활용해 에너지 절감

21) 일부 명사나 의존명사, 부사, 혹은 어간에 붙어서 서술하는 형태. 즉, 선행하는 명사나 부사의 의미 본질을 동작이나 상태로 나타내고자 할 때 선행하는 어휘에다가 ‘하다’를 붙여서 쓰는 형태로, “노래하다, 연구하다, 자랑하다, 사랑하다” 등이 이에 해당된다(백봉자, 1999; 한국어문법사전).

표 3. 헤드라인(HD기사)에서 단어의 출현빈도순

순서	단어	빈도	순서	단어	빈도	순서	단어	빈도
01	입주민	122	35	편익	7	68	투명성	4
02	아파트	116	36	환경	7	69	에너지	4
03	관리	62	37	만족도	7	70	현안	4
04	조성	42	38	각종	7	71	강화	4
05	주거환경	42	39	극대화	6	72	수립	4
06	쾌적	39	40	최선	6	73	마련	4
07	실현	34	41	이웃사랑	6	74	선정	4
08	투명	30	42	활성화	6	75	업그레이드	4
09	시설물	28	43	서비스	6	76	운동	4
10	철저	27	44	세심	6	77	전개	4
11	관리주체	26	45	유지보수	6	78	중진	4
12	절감	22	46	안락	5	79	공동체문화	3
13	관리비	22	47	홈페이지	5	80	입주민간	3
14	관리직원	21	48	대표회의	5	81	관심	3
15	화합	21	49	효율성	5	82	노후시설	3
16	공동체	19	50	체계적	5	83	부녀회	3
17	효율적	18	51	해결	5	84	가치	3
18	한마음	16	52	노력	5	85	절약	3
19	실시	16	53	합리적	5	86	난방방식	3
20	관리업무	15	54	의견	5	87	모범	3
21	개선	13	55	형성	5	88	주거만족도	3
22	운영	12	56	보수	5	89	점검	3
23	다양	12	57	마을	5	90	변경	3
24	단지	11	58	적극적	4	91	주민화합	3
25	활용	10	59	관리계획	4	92	공간	3
26	유지관리	10	60	공사	4	93	설치	3
27	참여	9	61	관리운영	4	94	향상	3
28	실천	9	62	주거	4	95	규정	3
29	안전	9	63	확보	4	96	주거공간	3
30	문화	9	64	수립	4	97	계획관리	3
31	추진	8	65	반영	4	98	활동	3
32	신뢰	8	66	구축	4	99	협력	3
33	아름	8	67	앞장	4	100	도입	3
34	편익증진	7						

작이나 상태로 나타내고자 할 때, 그 목적 대상으로 하는 내용을 살펴보자. 우선, ‘조성’의 경우를 보면(~을 조성(하다)의 형태), 문서에서 ‘(쾌적한) 주거환경’이나 ‘(살기 좋은) 아파트’가 대부분을 차지하고 있으며, 일부는 ‘공동체 문화’나 또는 ‘(아름다운) 환경’이 해당된다. 즉, 단지의 쾌적한 주거환경이나 또는 살기 좋은 아파트 조성이 모범적인 모습(image)으로 묘사되고 있다.

다음으로 ‘실현’의 경우에는 두 가지로 나누어서 알아볼 수 있는데, 하나는 ‘~관리+실현’의 형태이다. 여기서 ‘실현’의 목적어는 ‘관리’이고, 이를 꾸미는 형용어구는 대부분이 ‘투명한’이며, 일부 ‘효율적’이라는 어구가 등장하였다. 여기서 ‘투명한’이라는 형용사가 의외로 빈출하고 있는 점에 주목할 필요가 있다. 즉, ‘투명한 관리실현’이 강조되고 있는데, 이는 공동주택 운영이 체계적

이며 합리적으로 구성되어 투명하고 적합한 관리가 요구되고 있는 현실적 상황을 감안하면, 한편으로는 역설적으로 그러하지 못한 많은 아파트단지들과 대비된다는 의미에서 모범적인 사례로 선택되었다는 점에서 흥미롭다. 다른 하나는 ‘명사+실현’의 형태인데, 여기서 목적내용에 해당하는 명사는 “아파트공동체, 주민복지, 자치문화, 관리운영, 관리문화, 책임관리, 주거문화”의 단어들이다.

3) 네트워크분석

앞서 설명한 바와 같이 헤드라인(HD기사)에 대하여 공기단어간의 출현빈도 행렬을 만들어 네트워크분석을 이용하여 그 관계를 그림으로 나타내면 그림 4(a)와 같다. <그림 4>는 사회네트워크분석에서 가장 보편적인 Ucinet6.1²²⁾과 부속 지도화 프로그램인 NetDraw2.3을 이용하여 단어 간의 관계를 나타낸 것이다. 네트워크는 구성 요소인 노드(node, 여기서는 단어)들과 이들을 연결시켜주는 링크(link 또는 tie)들로 구성되는데, 여기서 공기단어 간의 긴밀 정도는 바로 출현빈도가 높을수록 강한 관계를 의미하는 것으로 그림에서는 굵기의 정도로 표시하였다.

그림에서 보는 바와 “입주민, 아파트, 관리, 조성, 주거환경”과 같은 단어들이 화면 중앙을 중심으로 배치되어 강하게 연결되어 있다. 말하자면, 개별 헤드라인 기사에서 이들 단어들의 조합에 의한 표현내용이 많다는 것을 말한다. 특히, 그림 4(a)의 화면 가운데에 위치한 ‘입주민’의 경우에는, “(다양한)의견, (편익)증진, (주거/입주민)만족도, 신뢰, 안전” 등의 공기단어와도 빈출하여, 결과적으로 입주민의 의견, 편익증진, 안전성확보, 주거만족도, 신뢰감 형성이 중요한 논점임을 나타내고 있다.

한편, 네트워크분석에서는 구성요소인 노드의 중심성(Centrality)을 비교할 수 있다. 네트워크 중심성은 연결망 체계에서 중심에 위치할수록 다른 구성 노드에 대하여 발휘할 수 있는 통제권 혹은 영향력이 상대적으로 크다는 것을 말한다(이우권, 1998). 공기단어 분석에 이를 유추하면, 단어와의 관계에서 중심성이 큰 단어일수록 영향력이 큰 단어 또는 일반적인 단어라고 할 수 있다. 표4는 공기단어 네트워크상에서 프리만(Freeman)의 중심성 상위 20개 단어를 추출하여, 전체 네트워크상에서 차지하는 비율을 나타낸 것이다. <표 4>는 공기단어의 중심성 점유 비율을 나타낸 것인데, 이 중에서 (a)는 헤드라인에서의 중심성으로, “입주민, 아파트, 관리”가 헤드라인 기사에서 중심적인 역할을 하고 있다. 중심성의 순위와 단어의 출현빈도 순위와를 비교할 때, 헤드라인에서는 순위에 큰 차이가 없으나, 중간제목과 본문기사에서는 다소 있다.

2. 중간제목(ST기사) 분석

1) 출현빈도 분석

앞서 헤드라인 분석에서 적용한 방법과 마찬가지로 816

22) (주)Analytic Technologies: <http://www.analytictech.com/>

표 4. 상위 20개 공기관어 중심성의 점유비율

(a)		(b)		(c)	
헤드라인(HD기사)		중간제목(ST기사)		본문 기사(FT기사)	
단어	비율	단어	비율	단어	비율
입주민(1)	10.5	입주민(1)	6.7	입주민(1)	10.7
아파트(2)	8.4	관리(2)	5.3	단지(2)	8.9
관리(3)	5.5	관리비(5)	4.5	입주자대표회의(5)	3.9
조성(4)	4.1	절감(3)	4.4	부녀회(4)	3.9
주거환경(5)	3.8	조성(7)	3.6	관리(3)	3.8
쾌적(6)	3.6	부녀회(6)	3.3	관리주체(7)	3.5
실현(7)	3.0	쾌적(13)	3.2	관리직원(6)	3.3
투명(8)	3.0	주거환경(9)	3.1	공사(8)	2.7
시설물(9)	2.7	투명(8)	2.9	관리사무소(10)	2.6
철저(10)	2.5	노력(10)	2.9	관리비(9)	2.3
절감(12)	2.3	공동체(12)	2.9	시설물(13)	2.2
관리비(13)	2.2	화합(11)	2.8	세대(11)	2.1
관리주체(11)	2.2	활동(15)	2.8	직원(14)	1.7
화합(15)	1.7	아파트(4)	2.2	관리사무소장(18)	1.6
공동체(16)	1.7	활발(18)	2.2	주거환경(15)	1.5
관리직원(14)	1.7	철저(14)	1.7	대표회의(12)	1.5
효율적(17)	1.6	단지(16)	1.6	운영(16)	1.4
한마음(18)	1.5	시설물(19)	1.6	입주(17)	1.3
실시(19)	1.2	효율적(31)	1.2	보수(21)	1.3
개선(21)	1.1	운영(17)	1.2	입주민들(24)	1.3
합계	64.3		60.1		61.5

주: 프리만(Freeman)의 연결정도(degree) 중심성(Centrality)으로서, 괄호 내는 각 문서에서의 출현빈도의 순위를 나타냄. 또한 비율은 전체 네트워킹상의 100개 단어 중에서 차지하는 비율임.

이 높은 순위를 보였다.

우선 ‘절감’의 목적어는 대부분 ‘관리비’로서, “~에 의한 관리비 절감”의 형태로 기술되었는데, 부분적으로 “노력, 극대화, 앞장”의 단어들이 선택적으로 부가되어 문장을 구성하고 있다. 말하자면, 중간제목으로는 ‘관리비 절감’이 중요한 논점이 되고 있다. ‘부녀회’의 경우에는 ‘입주자대표회의’와 함께 단지내 봉사활동을 포함한 다양한 부녀회 활동을 소개하기 위하여 중간제목으로 상당량 뽐히고 있다. ‘노력’의 경우에는 ‘주거환경 개선’이나 ‘투명한 관리’ 등이 목적어로 등장한다. 그리고 ‘조성’의 경우는 앞서 헤드라인과 유사한 내용이 목적어로 등장한다.

한편, 헤드라인 문서와 중간제목 문서에서 k=100 상위 출현빈도 단어 중에서 공통으로 출현하는 단어는 <표 6>에서 보는 바와 같이 총 63개이다. <표 6>에서 두 문서에서 차지하는 출현 순위가 크게 달라진 단어는 “부녀회, 활동, 관리주체, 노력, 절감, 관리비” 등이다. 헤드라인에서는 “쾌적, 실현, 시설물, 관리주체, 관리직원”의 단어가 중심어인데 비하여 중간제목에서는 “절감, 관리비, 부녀회, 노력”의 단어가 상대적으로 중심어를 이루고 있다.

2) 네트워크분석

공기단어 네트워크분석의 결과는 그림 4(b)에 나타나 있다. 중심어를 추출하여 보면 관리비(절감, 노력, 자체공사), 입주민(편의, 편익, 안전, 참여), 관리(시설물, 효율적,

표 5. 중간제목(ST기사)에서 단어의 출현빈도순

순서	단어	빈도	순서	단어	빈도	순서	단어	빈도
01	입주민	100	35	자체	12	68	주민의견	5
02	관리	65	36	관리직원	11	69	문화	5
03	절감	52	37	대표회의	11	70	이웃	5
04	아파트	51	38	관리운영	11	71	제공	5
05	관리비	49	39	활성화	11	72	증진	5
06	부녀회	46	40	유지관리	10	73	의견수렴	5
07	조성	41	41	최선	10	74	적극적	5
08	투명	40	42	관리서비스	10	75	관리주체	5
09	주거환경	40	43	봉사	9	76	도입	5
10	노력	37	44	합리적	9	77	아름	5
11	화합	35	45	설치	9	78	개선사업	5
12	공동체	32	46	앞장	9	79	전환	4
13	쾌적	29	47	참여	9	80	제고	4
14	철저	29	48	서비스	9	81	극대화	4
15	활동	28	49	환경	9	82	편의시설	4
16	단지	27	50	다양	9	83	제도	4
17	운영	25	51	자체공사	8	84	관리실현	4
18	활발	24	52	편익증진	8	85	예방	4
19	시설물	20	53	잠수입	7	86	수렴	4
20	실시	20	54	편익	7	87	강화	4
21	실현	18	55	입대	7	88	보수공사	4
22	활용	16	56	구축	7	89	에너지절약	4
23	실천	16	57	에너지	6	90	행정	4
24	주민	15	58	관리사무소	6	91	인력	4
25	개선	15	59	마련	6	92	CCTV	4
26	해결	15	60	편의	6	93	직원	4
27	입주자대표회의	15	61	책임관리	6	94	조경관리	4
28	공사	15	62	사랑	6	95	수립	4
29	봉사활동	14	63	안전관리	6	96	도모	4
30	안전	14	64	역점	5	97	투명성	4
31	효율적	13	65	절약	5	98	효율성	3
32	관리업무	13	66	각종	5	99	유지보수	3
33	세심	13	67	명확	5	100	리모델링	3

실현), 투명(관리운영, 관리실현), 조성(주거환경, 쾌적, 단지) 등이다(괄호 내는 대응하는 단어들 나타냄). 이를 요약해 보면, “입주민에 대한 편의성 증진과 주민참여에 의한 공동체 활성화, 시설물의 효율적 관리, 그리고 이 과정에서 투명한 관리운영과 실현을 통하여 궁극적으로 쾌적한 단지의 주거환경 조성”으로 논점이 정리된다.

3. 본문 기사(FT기사) 분석

1) 출현빈도 분석

<표 7>은 본문 기사에서 추출한 단어로 키워드만을 선별적으로 추출한 빈도순으로 정리한 것이다. 이것은 앞서 헤드라인과 중간제목 문서에서 선별과정 없이 출현빈도순으로 정리한 것과는 대조적이다. 이렇게 본문 기사에서는 선별적으로 키워드를 추출한 이유는, “지난해, 최근, 경우”와 같이 문서에서 출현빈도는 높으나 키워드로

표 6. 헤드라인 및 중간제목 문서의 출현 상위 100개 단어 중 63개 공통단어(헤드라인 문서에서의 순위, 중간제목 문서에서의 순위)

• 입주민(1, 1)	• 다양(23, 50)	• 효율적(49, 98)
• 아파트(2, 4)	• 단지(24, 16)	• 해결(51, 26)
• 관리(3, 2)	• 활용(25, 22)	• 노력(52, 10)
• 조성(4, 7)	• 유지관리(26, 40)	• 합리적(53, 44)
• 주거환경(5, 9)	• 참여(27, 47)	• 적극적(58, 74)
• 목적(6, 13)	• 실천(28, 23)	• 공사(60, 28)
• 실현(7, 21)	• 안전(29, 30)	• 관리운영(61, 38)
• 투명(8, 8)	• 문화(30, 69)	• 수립(64, 95)
• 시설물(9, 19)	• 하천(33, 77)	• 구축(66, 56)
• 철거(10, 14)	• 편익증진(34, 52)	• 앞장(67, 46)
• 관리주체(11, 75)	• 편익(35, 54)	• 투명성(68, 97)
• 절감(12, 3)	• 환경(36, 49)	• 에너지(69, 57)
• 관리비(13, 5)	• 각종(38, 66)	• 강화(71, 87)
• 관리직원(14, 36)	• 극대화(39, 81)	• 수립(72, 86)
• 화합(15, 11)	• 최선(40, 41)	• 마련(73, 59)
• 공동체(16, 12)	• 이웃사랑(41, 34)	• 증진(78, 72)
• 효율적(17, 31)	• 활성화(42, 39)	• 분회(83, 6)
• 실시(19, 20)	• 서비스(43, 48)	• 절약(85, 65)
• 관리업무(20, 32)	• 세심(44, 33)	• 설치(93, 45)
• 개선(21, 25)	• 유지보수(45, 99)	• 활동(98, 15)
• 운영(22, 17)	• 대표회의(48, 37)	• 도입(100, 76)

※음영처리된 부분은 출현 순위가 두 배 이상 차이가 난 경우

서의 역할이 크게 없는 단어²⁴⁾가 많기 때문이다. 그래서 앞서 설명한 바와 같이 아파트관리 분야에서 중요성과 대표성이 있다고 판단되는 단어만을 k=100에서 추출·선택하였다(표 7). 여기서 추출하는 단어 k의 개수에 대해서는 특별히 정해진 규칙은 없으므로 분석의 목적에 따라 적절히 선택하면 될 것이나, 단어의 출현빈도는 거의 지수분포(exponential distribution)를 띄므로 단어의 구성분포를 고려하여 상위 k까지 추출하면 된다. 이미 헤드라인과 중간제목의 경우에도 k=100으로 설정하였다. 일반적으로 문서에서 어휘의 분포에는 ‘지프의 법칙’(zip’s law)²⁵⁾이 성립하는 것으로 알려져 있다.

<표 7>에서 상위 10위까지 단어를 헤드라인과 중간제목의 경우와 비교하여 보면, ‘입주자대표회의’, ‘관리직원’, ‘관리주체’, ‘관리사무소’와 같은 단어가 눈에 띈다.

2) 네트워크분석

분석결과는 그림 (4c)에 표시되어 있다. 우선 헤드라인

24) 자연언어처리 분야에서는 일반적으로 이들 색인어로서 선택하지 않는 단어를 스톱 워드(stop word; 불용어)라고 부른다. 한글의 경우 조사, 접속사, 어미 등이 이에 해당하며, 영어에서는 a(an), the, that, this, in, but과 같은 단어가 해당된다. 그러나 본 연구에서는 이를 더 확장하여 일반 명사까지 적용하고 있다.

25) 지프의 법칙이란, 어휘의 사용빈도의 크기에 그 순위를 곱하여 얻는 수치는 거의 일정하다는 것이다(빈도×빈도순위 = 상수). 이러한 지프의 법칙에 따르는 현상으로는 유명한 도시인구의 분포, 유리조각의 크기 분포, 나아가 기업소득의 분포 등 경제현상에도 이 법칙이 적용된다고 알려져 있다.

표 7. 본문 기사(FT기사)에서 단어의 출현빈도순

순서	단어	빈도	순서	단어	빈도	순서	단어	빈도
01	입주민	3063	35	지하주차장	159	68	불우이웃	70
02	단지	2185	36	어린이	143	69	분리수거	70
03	관리	903	37	공간	139	70	반상회	69
04	부녀회	889	38	이웃	138	71	조경관리	69
05	입주자대표회의	832	39	서비스	132	72	하자보수	66
06	관리직원	711	40	쓰레기	129	73	입주민간	66
07	관리주체	708	41	동별	129	74	보일러	65
08	공사	702	42	에너지	119	75	관리동	62
09	관리비	581	43	유지관리	115	76	이웃간	62
10	관리사무소	470	44	동대표	115	77	배관	62
11	세대	461	45	홈페이지	114	78	지하	61
12	대표회의	455	46	하자	108	79	관리운영	61
13	시설물	442	47	화단	108	80	놀이터	60
14	직원	383	48	노인	107	81	반장	60
15	주거환경	364	49	노인회	106	82	현관	60
16	운영	333	50	대표회장	103	83	규정	58
17	입주	324	51	보수공사	100	84	재활용품	58
18	관리사무소장	262	52	잡수입	98	85	출입구	57
19	주민	258	53	청소	96	86	자전거	57
20	관리소	246	54	노후화	92	87	안전관리	57
21	보수	245	55	봉사활동	90	88	나무	56
22	관리업무	244	56	어린이놀이터	89	89	바자회	54
23	승강기	233	57	부녀회장	88	90	외벽	54
24	입주민들	224	58	난방	87	91	한마음	53
25	안전	198	59	전기	86	92	장기수선 계획	53
26	관리소장	192	60	수익금	86	93	공용부분	52
27	공동체	189	61	부녀회원	82	94	사고	52
28	CCTV	186	62	교체공사	81	95	범죄	51
29	민원	184	63	경비원	80	96	안전사고	51
30	환경	174	64	관리서비스	75	97	독거노인	51
31	수목	167	65	차량	75	98	옥상	50
32	소장	163	66	노인정	72	99	우수관리	50
33	식재	161	67	이웃사랑	71	100	주차장	49
34	행사	161						

이나 중간제목과는 달리 공기단어 네트워크상의 링크가 상호 긴밀하게 연결되어 전체적으로 복잡한 형상이다. 화면의 중앙에는 “입주민, 단지, 관리, 입주자회의, 관리직원, 관리사무소, 관리비, 시설물, 공사, 부녀회” 등의 단어가 배치되어, 본문 기사에서는 이들 단어가 중심성을 이루고 있음을 알 수 있다.

흥미롭게도, 공기단어의 배치 관계에 주목하여 자세히 살펴보면 공동주택관리에서 주요한 세 가지 관리 개념을 읽을 수 있다. 바로 유지관리, 운영관리, 생활관리가 그것인데, “입주민, 단지”를 중심으로 화면의 우측에는 “승강기, 보일러, 지하주차장, 노후화, 보수공사” 등을 구성 단어로 하는 ‘A.유지관리’의 개념이 대응하고, 중앙 상단에는 “관리비, 입주자대표회의, 관리주체, 관리사무소” 등

을 구성 단어로 하는 ‘B.운영관리’의 개념이 대응하고 있다. 그리고 화면 좌측 하단에는 “이웃사랑, 봉사활동, 노인회, 기금, 행사, 문화, 공동체” 등을 구성 단어로 하는 ‘C.생활관리’의 개념이 대응하고 있다.

3) 군집분석

이번에는 2장3절에서 설명한 바와 같이 공기단어 간의 출현빈도를 군집적으로 하여 단어들을 군집화(grouping)함으로써 위계관계에 따른 문서의 의미를 고찰하였다. <그림 5>는 공기 출현빈도가 낮은 단어부터(E의 독거노인, 우수관리, 주차장 등) 군집화하여 가장 빈도수가 높은 ‘입주민’까지 100개의 키워드를 그룹핑한 것이다. 그림에서 알 수 있듯이 상위 A에서 하위 E로 갈수록 군집화된 그룹이 많고 복잡해지는데, 이는 앞서 설명한 바와 같이 하단으로 갈수록 공기 출현빈도가 높아지기 때문에 기인하는 것이다. 따라서 해석의 편의상 상위수준A에서부터 하위수준 E까지 군집화된 그룹의 특성을 간략히 살펴보면, 본문 기사 전체를 통틀어 ‘입주민’과 ‘단지’가 중심적 어구임을 알 수 있다.

그룹A에서는 “관리사무소, 입주자대표회의, 부녀회”로 대표되는 관리주체가 등장하고 있다. 다음으로 그룹B에서는 “보수, 승강기, CCTV”와 같은 유지관리에 관한 부분이 많이 차지하며, 그룹C에서는 “쓰레기, 행사, 노인회, 봉사활동”의 생활관리에 관한 내용이 많다. 바로 차하위

그룹D에서는 “청소, 보일러, 차량, 놀이터” 등의 유지관리가 많으며, 마지막 그룹E에서는 주로 유지관리와 관련 깊은 항목이 나열되어 있다.

한편, <그림 5>의 좌측부분은 군집화된 그룹의 크기가 우측과 비교하여 상대적으로 작아 별도의 기호로는 표시하지 않았는데, 대부분의 내용은 생활관리와 유지관리에 관한 것이다. 이렇게 본다면 운영관리에 관한 내용은 그다지 눈에 띄지 않아 대조를 보일뿐만 아니라, 앞서 헤드라인이나 중간제목의 출현빈도에서 ‘투명한 관리’가 우위를 점하여 운영관리의 내용이 부각되었던 것과는 상반된 결과를 보여준다. 그런데 한편으로 프로그래밍 코드에 의하지 않고 실제로 본문 기사를 내용분석해 보면, 많지는 않지만 ‘운영관리’에 대해서도 제법 서술하고 있다. 이렇게 상반된 이유는, 본문 기사에서는 헤드라인이나 중간제목 문서와는 달리 ‘투명’과 같은 운영관리와 관련 깊은 단어를 키워드에서 배제하였다는 점에 부분적으로 기인하는 것으로 생각된다. 그러나 한편으로는 취재기사가 헤드라인이나 중간제목에서 비록 ‘투명한 관리 실현’의 형태로 운영관리에 대한 내용을 핵심어로 뽑으면서도 실제의 본문내용에서는 운영관리에 관한 구체적인 언급이 미흡했기 때문인 것으로 풀이된다. 또한 일반적으로 운영관리의 내용은 계획적인 내용이 많아 유지관리나 생활관리와 같은 실천적인 내용과 이를 표현하는 구

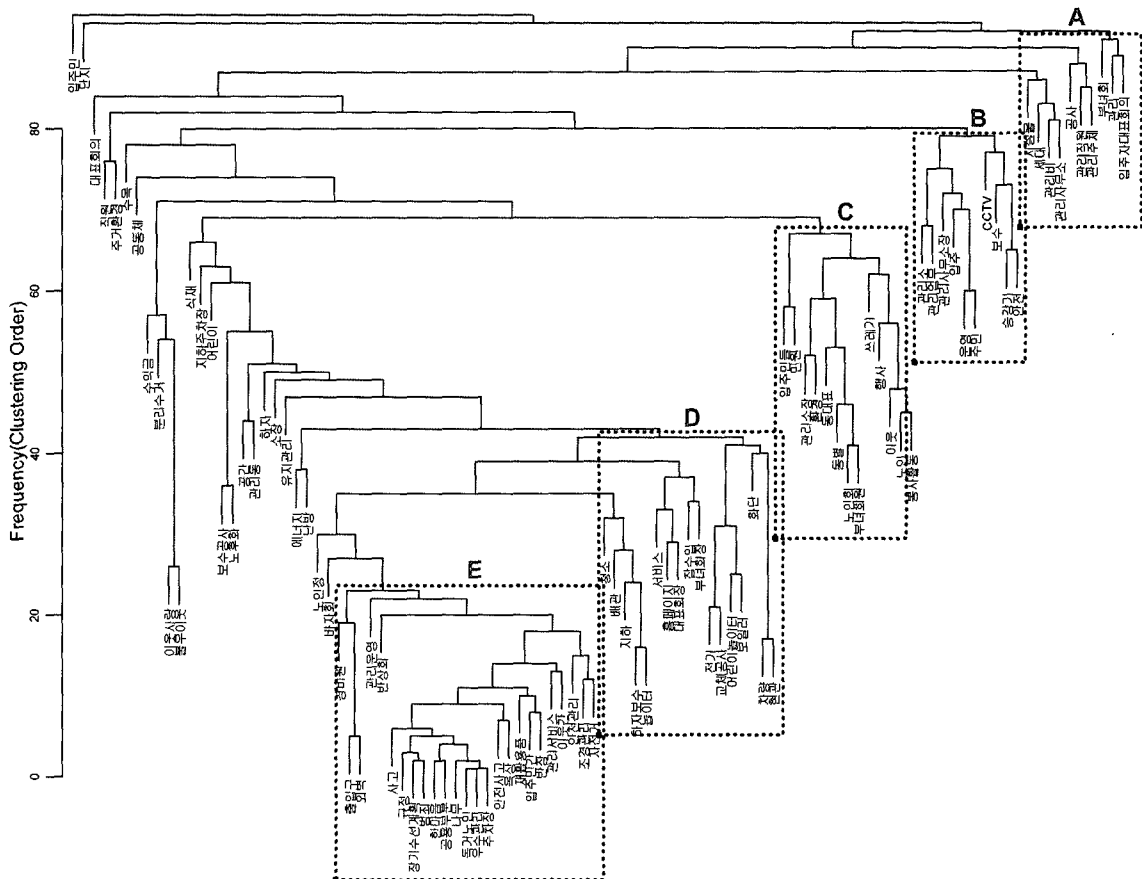


그림 5. 공기단어 행렬에 의한 단어의 군집분석(k=100)

체성을 갖춘 단어로 표현되지 않기 때문에 기인하는 것으로 해석된다.

VI. 결 론

본 연구는 최근 매스컴의 우수 아파트단지 사례에서 소개하고 있는 모범적인 아파트 관리에 관한 내용을 언어학적 특성에 주목하여 정량적으로 분석하여 보았다. 즉, 기존의 내용분석 방법과는 달리 텍스트마이닝의 자연언어처리 기법을 응용하여 특정 문서에서 언어의 통계량을 계산하여 이를 통계학 및 사회학적 분석기법과 연계함으로써, 문서에 잠재되어 있는 전달 메시지를 논점의 형태로 파악하여 본 이 분야에서의 실험적인 연구이다. 아파트관리신문에서 1999년 12월부터 2006년 3월 현재까지 연재하고 있는 탐방기사 282개를 텍스트형식으로 추출하여 헤드라인, 중간제목, 본문 기사의 세 가지 형태로 분류하여 언어통계학적 특성을 고찰하였는데, 일련의 분석 과정에서 얻은 결과를 간단히 정리하면 다음과 같다.

첫째, 전달 희망 내용을 핵심적으로 축약한 헤드라인 기사에는 “입주민, 아파트, 관리, 조성”의 빈도순으로 단어들 이 출현하였으며, “조성”의 목적내용은 “단지의 쾌적한 주거환경이나 또는 살기 좋은 아파트”이었다. 또한 ‘실현’의 경우에는 ‘투명한 관리’와 ‘아파트공동체’가 실현의 주요 목적내용을 이루고 있다.

둘째, 문서에서 가장 빈출하는 ‘입주민’의 경우에는 입주민의 “의견, 편익증진, 안전성확보, 주거만족도, 신뢰감형성”이 주요 이슈이다. 헤드라인과 중간제목에서 각각 출현하는 상위 100개 단어 중에서 공통으로 출현하는 단어는 63개인데, 이 중에서 출현 순위에서 큰 변동을 보인 단어로는 “관리비, 부녀회, 공사” 등이었다. 이들 단어는 헤드라인보다는 중간제목에서 기술됨으로써 상대적으로 비중 있게 다루어지는 구체적인 논점이다.

셋째, 816개의 중간제목에서는 ‘관리비 절감’이 중요한 논점이었으며, ‘주거환경의 개선’이나 ‘투명한 관리’가 ‘노력’의 목적 내용으로 등장하였다. 네트워크분석으로 종합하였더니 “입주민에 대한 편의성 증진과 주민참여에 의한 공동체 활성화, 시설물의 효율적 관리, 그리고 이 과정에서 투명한 관리운영과 실현을 통한 쾌적한 주거환경의 단지 조성”이라는 논점으로 요약되었다.

넷째, 본문 기사에서는 네트워크분석의 배치 관계에서 유지관리, 운영관리, 생활관리라는 관리에서의 세 가지 주요 개념이 “입주민, 단지”를 중심으로 배치되어 나타났다. 유지관리에서는 “승강기, 보일러, 지하주차장, 노후화, 보수공사” 등을 중심으로, 운영관리에서는 “관리비, 입주자대표회의, 관리주체, 관리사무소”를 중심으로, 그리고 생활관리에서는 “이웃사랑, 봉사활동, 노인회, 기금, 행사, 문화, 공동체”를 중심으로 각각의 개념이 표출되었다.

다섯째, 공기단어를 군집화하였더니 대부분 생활관리와 유지관리에 관한 내용이었고, 운영관리에 관한 것은 상

대적으로 적었다. 헤드라인이나 중간제목에서는 ‘투명한 관리 실현’의 형태로 운영관리에 관한 내용을 핵심어로 뽑으면서도 실제의 본문내용에서는 운영관리에 관한 언급이 미흡하여 흥미로운 사회언어학적 특성을 보였다.

이상과 같은 결과는 살기 좋은 우수 아파트단지의 조성 조건이자 모범적 이미지로서, 우리가 가지는 일반적 이미지와 합치한다. 즉, 투명한 운영관리, 입주민의 화합과 참여에 의한 공동체 실현, 이를 통한 쾌적한 주거환경의 조성인데, 이는 동시에 향후 아파트관리의 방향과 관리체계 구축에서 주요하게 다루어져야 할 내용임을 시사한다. 무엇보다도 본 연구에서는 이를 계량적으로 입증할 수 있었다는 데에 큰 의의가 있다고 할 수 있다.

한편, 연구의 한계로는 첫째, 분석대상으로 하고 있는 (주)아파트관리신문사의 탐방기사가 사회일반에서 얼마나 보편적인 대표성을 갖고 있는지는 점이다. 둘째, 아울러 취재내용이 취재기자나 회사의 취재방침에 따라 자의든 타의든 왜곡될 수 있다는 점이다. 이는 언론학에서 말하는 ‘의제설정기능’(agenda-setting function)을 의미하는 것으로, 매스미디어가 특정 논점을 강조하여 다른 논점을 경시하는 등 일정한 우선순위를 매김으로써 수용자의 현실적 인지에 커다란 영향을 미친다는 점이다. 셋째, 분석기법상의 한계로 인간의 미묘한 언어표현과 언외의 의미론적(semantic) 특성에 의한 메시지를, 단순히 문서에서의 출현빈도 등 통사론적(syntactic) 방법으로 과연 얼마나 추출할 수 있는지는 문제이다.

이러한 한계에도 불구하고 적용의 가능성과 잠재적인 유용성은 높다고 지적되고 있으며, 또한 한계를 극복하기 위한 학제적인 연구가 활발한 만큼 주거학 분야에서 도 관심을 가지고 적용사례를 축적할 필요가 있을 것이다.

참 고 문 헌

1. 김정엽 외(2004), 의사결정 트리 기법을 이용한 인천광역시 도시 성장 모델링, 국토계획 39(4), pp.31-44.
2. 김태운·이창무(2005), 임차인의 임대계약 선택에 있어서 데이터마이닝 기법들을 이용한 비교분석, 대한국토도시계획학회 정기학술대회 논문집, pp.679-692.
3. 박우열 외(2005), 서포트 벡터 머신을 이용한 개산견적 평가모델에 관한 연구, 대한건축학회 논문집(구조계) 21(12), pp.191-198.
4. 박철수(2005), 대중소설에 나타난 아파트의 이미지 변화 과정 연구, 대한건축학회 논문집(계획계) 21(1), pp.189-200.
5. 박한우·Loet Leydesdorff(2004), 한국어의 내용분석을 위한 KrKwic 프로그램의 이해와 적용, Journal of The Korean Data Analysis Society 6(5), pp.1377-1388.
6. 백봉자(1999), 한국어문법사전, 연세대 출판부.
7. 손세관·김승언(2000), 1990년대 신문광고에 나타난 우리나라 아파트의 계획적 특성에 관한 내용분석, 대한건축학회 논문집(계획계) 16(11), pp.105-112.
8. 신용재(1996), 아파트 분양광고문에 나타난 입지성 - 부산권 신문광고를 중심으로, 대한건축학회 논문집 12(7), pp.61-71..

9. 신창현(2003), 인쇄매체 광고를 통해 본 우리나라 주거문화의 경향에 대한 고찰 신창현, 서울대학교 대학원 건축학과 석사.
10. 이상우(1982), 헤드라인의 기능과 실제 한국 신문들이 안고 있는 문제점을 중심으로, 언론중재 겨울호.
11. 이성미·이연숙(2004), 주요일간지의 아파트 관련기사에 나타난 주거가치 특성에 관한 연구, 한국실내디자인학회 학술발표대회 논문집 6(1), pp.91-94.
12. 이우권(1998), 연결망분석의 행정학적 함의, 전북행정학회 동계 학술발표회.
13. 이재혁·제해성(2005), 신문광고와 소비자 의식조사 비교에서 나타난 웰빙아파트의 건축계획적 특성 연구, 대한건축학회 논문집(계획계) 21(12), pp.127-134.
14. 이종국(2004), 건설데이터 웨어하우스 시스템 프로토타입 기초 연구, 한국건설관리학회 논문집 5(3), pp.166-173.
15. 홍윤영 외(2004), 아파트 신문광고에 나타난 사회적 차별성에 대한 연구, 대한건축학회 논문집(계획계) 20(11), pp.3-10.
16. 姜涌 ほか(1999), 1950~1970年代の中國における建築雑誌に現れる建築用語の統計的分析-中國建築の近代化過程における建築家の言説に関する研究 その1, 『日本建築學會計畫系論文集』 Vol.516.
17. 高橋由光 (2002), 「環境」概念の構造的關係 - 「環境」に関する記事の時系列的內容分析, 東京大學新領域創成科學研究科環境學專攻修士論文.
18. 松田芳雄 (2001), 顧客分析とデータマイニングの動向, 『UNISYS TECHNOLOGY REVIEW』 第68號.
19. 見田宗介 (1965), 「質的」なデータ分析の方法論的な諸問題, 『社會學評論』 15(4), 日本社會學會, pp.79-91.
20. 市村由美・鈴木優 (2001), テキストマイニング技術と應用, 『東芝レビュー』 56(5), pp.19-22.
21. 崔廷敏・淺見泰司 (2003), 大量の屬性データからの興味深いルール抽出手法: 新規分譲マンションの契約者データへの適用, 『日本建築學會計畫系論文集』 Vol.564, pp.303-310
22. 崔廷敏(2004), 住情報及び住意識に関する探索的分析, 東京大學工學系研究科都市工學專攻博士論文.
23. Adriaans, P & Zantinge, D. (1998) 『データマイニング』, 共立出版.
24. Bounsaythip, C & Rinta-Runsala, E. (2001), Overview of Data Mining for Customer Behavior Modeling, VTT INFORMATION TECHNOLOGY RESEARCH REPORT.
25. Choi, J.M. & Asami, Y. (2002), An Alternative Method for Extracting Unexpected Patterns from Huge Attributes Using Conditional Contingency Table in Marketing, A. Zanasi, C.A. Brebbia, N.F.F.E. Ebecken, P. Melli (eds.) Data Mining III, WIT Press, Southampton, pp.113-120.
26. Langley (1986), Editorial: Machine Learning and Discovery, Machine Learning, 1, pp.363-366.
27. Lebart, L., Salem, A., & Berry, L. (1997), Exploring Textual Data, Kluwer Academic Publishers.
28. Manning, C.D. & Schutze, H. (1999), Foundations of Statistical Natural Language Processing (Cambridge, MA, MIT Press).

(接受: 2006. 4. 20)

