

## 빅데이터를 활용한 다이어트 현황 및 네트워크 분석

정은진 · 장은재<sup>†</sup>  
동덕여자대학교 식품영양학과

## Tendency and Network Analysis of Diet Using Big Data

Eun-Jin Jung · Un-Jae Chang<sup>†</sup>

Dept. of Food &amp; Nutrition, Dongduk Women's University, Seoul 02748, Korea

## ABSTRACT

Limitation of a questionnaire survey which is widely used is time and money, limited numbers of participants, biased confidence interval and unreliable results. To overcome these, we performed tendency and network analysis of diet using big Data in Koreans. The keyword on diet were collected from the portal site Naver from January 1, 2015 until December 31, 2015 and collected data were analyzed by simple frequency analysis, N-gram analysis, keyword network analysis and seasonality analysis. The results showed that diet menu appeared most frequently by N-gram analysis, even though exercise had the highest frequency by simple frequency analysis. In addition, keyword network analysis were categorized into four groups: diet group, exercise group, commercial diet program company group and commercial diet food group. The analysis of seasonality showed that subjects' interests in diet had increased steadily since February, 2015, although subjects were most interested in diet in July, these results suggest that the best strategies for weight loss are based on diet menu and starting diet before July. As people are especially sensitive to diet trends, researches are needed about annual analysis of big data.

**Key words** : Big Data, tendency, questionnaire survey, network analysis, diet menu, exercise

## 서론

급속한 경제발전과 생활수준 향상으로 식생활은

다양해지고, 편리한 생활을 위해 발전한 자동화는 신체 활동량을 감소시켜 운동량 부족으로 비만 인구를 증가시키고 있는 추세이다. 또한 대중매체의 영향으로 건강과 미에 대한 관심이 높아지면서 건강한 신체와 아름다운 몸매를 위해 다이어트를 하는 사람들이 늘어나고 있다.

일반적으로 다이어트는 건강이나 미용을 목적으로 체중을 조절하기 위한 규정식 또는 식이요법으로 정의된다. 그러나 다이어트의 정의는 연구자에 따라 약간의 차이가 있는데 Herman & Polivy(1988)는 음식을 먹고자 하는 욕구를 충족시키지 않고 저

This research was supported by grants from Dongduk Women's University.

접수일 : 2016년 10월 6일, 수정일 : 2016년 10월 19일,

채택일 : 2016년 10월 19일

<sup>†</sup> Corresponding author : Un-Jae Chang, Department of Food and Nutrition, Dongduk Women's University, 23-1 Wolgok-dong, Seongbuk-gu, Seoul 02748, Korea

Tel : 82-2-940-4464, Fax : 82-2-940-4610

E-mail : uj@dongduk.ac.kr

ORCID : <http://orcid.org/0000-0002-9471-3380>

지하려고 인지하는 노력이라고 하였고, Shin(1997)은 체중감량을 위해 음식 섭취를 제한하기 위한 행동과 노력이라고 하였고, Turner(1996)는 개인의 신체와 몸매를 관리하고 유지하는 것이라고 정의했다. 즉, 다이어트는 식이를 규제하거나 제한함으로써 체중감량을 이끌고 신체를 관리하는 것이라고 볼 수 있다. 따라서 과거에 젊은 여성에게만 해당되었던 다이어트는 여성뿐만 아니라 남성, 청소년 등 다양한 계층으로 확산되고 있으며(Van 2007), 많은 사람들은 다이어트에 대해 관심을 가지고 시도하고 있다.

다양한 계층의 사람들은 다이어트를 시도하기 위해 손쉽고 빠르게 정보를 주고받을 수 있는 인터넷을 활용하여 정보를 얻고있다. 인터넷은 접근성이 용이하고 시간과 장소에 대한 제약이 없으며, 많은 정보를 포함하고 있어 정보 획득에 중요한 원천이 되고 있다(Koh & Yoon 2003; Chang 등 2004).

인터넷을 통해 전 세계 사람들과 교류가 가능해지고 다양하고 이색적인 다이어트 정보를 접할 수 있게 되어 사람들은 자신의 신체적 차이를 고려하여 자신에게 맞는 다이어트 방법을 검색하고 시도할 수 있게 되었다. 또한 인터넷을 이용해 다이어트 일기와 같은 문자의 형태뿐 아니라 사진, 동영상 등 다양한 형태로 정보를 주고받을 수 있게 되었다. 결국 모바일, 노트북 등과 같이 휴대가 용이한 기기의 사용 증가로 인해 다양한 형태의 다이어트 데이터 정보는 거대한 규모가 되었고, 데이터 처리속도와 생성속도 또한 빨라져 빅데이터(Big Data)의 형태를 갖추게 되었다.

세계경제포럼(World Economic Forum 2012)에서는 빅데이터를 떠오르는 10대 기술 중의 하나로 선정하였다. 이렇듯 빅데이터를 통해서 가치 중심적이고 경쟁력있는 정보를 다변화된 현대 사회에 제공할 수 있는 기술이기에 그 중요성이 꾸준히 증가하고 있다. 이를 증명하듯 빅데이터를 활용한 사례들이 많은데, 서울특별시에서 심야시간대에 안전하고 편리하게 시민이 귀가할 수 있도록 2013년부터 시행한 심야버스 운행은 빅데이터를 이용한 대표적인 사례

로 꼽을 수 있다. 서울특별시는 KT 통화데이터 30억건과 서울 교통데이터를 분석하여 8개의 심야버스 노선을 개설하였고, 이제는 하루 평균 7,500명의 시민이 이용하는 대중교통으로 자리매김하고 있다. 시행 2년이 지난 2015년에는 누적 이용자가 500만명이 넘어섰고 높은 이용률과 수요로 인해 운행횟수 증가와 버스 증차를 계획하고 있다(마경근 2015). 이처럼 빅데이터는 단순한 데이터의 집합이 아니라 유의미하고 경제적인 가치를 이끌어 나가는데 도움을 줄 수 있다.

빅데이터를 활용하는 또 다른 사례는 설문조사를 대신할 수 있다는 것이다. 연구자들은 필요한 정보를 수집하기 위해서 설문조사 방법을 사용하는데(Boynton & Greenhalgh 2004), 설문조사는 많은 비용과 시간이 필요하고, 모집단을 잘못 반영하면 추출된 표본이 모집단 전체를 대표한다고 간주할 수 없고, 편향된 신뢰구간, 표준오차 등 한계점으로 인해 잘못된 결과를 얻을 수 있다(Seol & Chung 2000). 따라서 본 연구는 이를 해결하고자 빅데이터를 통한 새로운 분석 방법을 시도하고 활용하고자 하였다.

이에 본 연구에서는 포털 사이트인 네이버의 2015년 1월 1일부터 2015년 12월 31일까지 1년간 빅데이터로 ‘다이어트’ 키워드가 포함된 문장을 수집한 후, 다이어트에 관련된 키워드가 무엇인지 살펴보는 빈도 분석, 다이어트 키워드와 관련 키워드들 간의 동시 출현한 빈도와 키워드 위치에 따라 방향성을 부여하는 N-gram 분석, 단어 간 연관성과 흐름을 파악하여 유사한 단어끼리 군집 형성을 분석하는 키워드 네트워크 분석 그리고 다이어트 키워드 월별 출현빈도를 통한 계절성 분석을 하여 다이어트에 관한 현황을 살펴보고 앞으로 어떻게 활용할 수 있는지 알아보하고자 한다.

## 연구방법

### 1. 연구 대상 및 기간

다이어트 현황 및 네트워크 분석을 위한 데이터를 수집하기 위해 본 연구에서는 국내에서 가장 인지도가 높은 인터넷 포털 사이트인 네이버를 분석 대상으로 선정하였고, 데이터 수집을 위한 키워드로 ‘다이어트’를 사용하였다. 다이어트는 시작하려는 날짜가 지정되어 있지 않고 개인마다 다이어트를 시작하는 시기가 다르기 때문에 한 해가 시작되는 1월부터 끝나는 12월까지 분석하고자 했다. 이에 본 연구에서는 2015년 1월 1일부터 2015년 12월 31일까지 1년간을 연구 기간으로 선정하였다.

### 2. 데이터 수집 및 전처리

데이터 수집을 위하여 Python 2.7(Pycon, USA) 프로그램을 이용하였고 네이버 검색 Application Programming Interface(API)의 기능을 이용하여 네이버 블로그, 웹문서, 뉴스 그리고 카페에서 ‘다이어트’ 키워드가 포함된 문장을 수집하였다. 수집된 문장은 은전한닢 프로젝트에서 최소한의 변경으로 한국어 특성에 맞게 형태소 분석을 해주는 Mecab 프로그램을 이용하여 다이어트에 연관된 명사를 추출하였다.

Mecab을 통해 명사 형태로 분리된 키워드는 데이터 분석을 위해 다음과 같은 전처리 작업을 수행하였다. 첫째, 같이 사용되는 단어로 2개 이상 분리된 단어들은 다시 하나의 단어 형태로 변환하였다. 예를 들어 ‘워킹맘’처럼 한 단어로 사용되는 단어가 ‘워킹’, ‘맘’으로 분리된 경우 다시 하나의 단어 형태인 ‘워킹맘’으로 변환하였다. 둘째, 단어의 형태가 약간 다르지만 동일한 뜻으로 사용되는 단어들의 경우에는 하나의 단어로 통일하였다. 예를 들어 ‘여성’, ‘여자’는 같은 뜻으로 간주하여 ‘여성’으로 통일하여 변환하였다. 마지막으로 ‘가’, ‘이’, ‘들’ 등 조사, 대명사 등으로 분리되어서 분석에 사용할 수 없

는 형태는 삭제하였다. 이에 ‘다이어트’ 키워드를 포함한 연관 키워드가 총 108,543개 나타났으며, 총 단어의 빈도수는 13,208,746개가 도출되었다.

### 3. 분석 방법

#### 1) 단순빈도 분석

수집된 방대한 양의 데이터 내용을 직관적으로 이해하기 위해서(Lee 등 2014)는 주제 키워드와 연관된 키워드들의 출현빈도를 높은 순서대로 나열하여 분석하면 유용하다. 연관된 키워드들은 주제 키워드에 관심과 흥미를 반영한 키워드로 간주할 수 있기 때문이다. 따라서 본 연구에서는 수집된 데이터에서 출현빈도가 높은 고빈도 키워드들이 무엇인지 단순빈도 분석을 통해 분석하였다.

#### 2) N-gram 분석

단순빈도 분석을 통해서 데이터 전체의 빈도를 분석할 수 있지만 주제 키워드와 연관 키워드 간의 동시 출현 및 밀집정도를 확인할 수 없기 때문에 본 연구에서는 N-gram 분석을 시행하였다. N-gram 분석은 연구자가 지정한 n개의 어절 또는 음절을 연속적으로 분류한 후 단어들 간의 밀집정도를 측정하고 빈도를 분석하였다(Hwang 등 2012).

예를 들어 ‘다이어트 성공의 지름길은 식단조절이다’라는 문장을 N-gram을 통해 분석하고자 할 때 n=2이면 음절 단위로 [‘다’, ‘이’], [‘이’, ‘어’], [‘어’, ‘트’], [‘트’, ‘ ’], [‘ ’, ‘성’], ..., [‘절’, ‘이’], [‘이’, ‘다’]로 분석한다. 만약 n=3이면 어절 단위로 [‘다이어트’, ‘성공의’, ‘지름길은’], [‘성공의’, ‘지름길은’, ‘식단조절이다’], [‘지름길은’, ‘식단조절이다’, ‘ ’], [‘식단조절이다’, ‘ ’, ‘ ’]로 분석한다.

따라서 본 연구에서는 키워드단위로 분리하여 N-gram 분석을 실시하였다. 또한 N-gram의 keyword1과 keyword2는 방향성을 나타내는 것이므로 이를 활용하여 키워드 간의 방향성도 확인하였다.

3) 키워드 네트워크 분석

키워드 네트워크 분석은 문장에서 명사, 형용사 형태의 키워드들 간의 연관성 관계를 파악하고 분석하여 키워드 간의 연결관계를 연결망 형식으로 추출할 수 있도록 한 방법이다(Diesner & Carley 2004; Ahn 2012). 네트워크는 개체(actors)로 나타내는 노드(node)와 관계를 나타내는 링크로 구성되어 있으며(Moon 2013), 본 연구에서는 키워드가 노드를 나타내고 키워드와 키워드의 연결이 링크로 나타난다. 따라서 키워드 네트워크 분석을 통해 전체적인 네트워크 구조, 링크와 노드들의 형태 및 특성, 노드들의 방향성 및 관계 등을 시각적으로 표현할 수 있다(Ahn & Oh 2015). 키워드 네트워크 분석을 위해 데이터를 소셜 매트릭스 형태로 구현해야 하는데 이는 (주)더아이엠에서 제공하는 소셜 매트릭스 프로그램인 '텍스톰(Textom)'을 활용하였다. 키워드 네트워크 분석 중 CONCOR 분석을 이용하였고, 이를 시각적으로 나타내기 위해 Ucinet 6 프로그램 (Analytic Technologies, USA)을 이용하여 시각화하였다.

(1) CONvergence of iterated CORrelations(CONCOR) 분석

CONCOR 분석은 동시에 출현한 단어들을 매트릭스의 피어슨 상관관계에 따라 노드들의 블록을 식별하고 블록들 간의 관계를 파악하여(Wasserman &

Faust 1994) 유사성을 지닌 키워드들 간의 관계를 이루고 군집을 형성하는 분석이다.

4) 계절성 분석

다이어트에 대해 관심과 흥미가 많아지면 블로그, 카페 등에 다이어트 키워드 자체의 노출 횟수가 많아질 것으로 간주하여 다이어트의 계절성 분석을 위해 다이어트 키워드를 월별 출현빈도로 분석하였다.

결 과

1. 단순빈도 분석

다이어트 키워드와 관련된 키워드가 무엇인지 살펴보기 위하여 키워드 출현빈도를 알아보았다. Table 1은 다이어트 키워드 검색 시 연관되어 출현한 키워드 중 상위 20위까지 선택하여 나타낸 빈도표이다. 분석결과 다이어트와 관련하여 출현빈도가 높은 키워드는 운동(178,197개), 건강(80,173개), 식단(79,675개), 효과(79,604개) 등의 순으로 나타났다. 이에 다이어트와 관련된 키워드로 운동, 건강, 식단이 빈번히 언급되는 것을 확인할 수 있었다. 또한 상위 30위 안에는 점심(37,549개), 여름(31,051개) 등을 확인할 수 있었다.

2. N-gram 분석

키워드와 키워드 간의 밀집정도를 통해 동시 출현빈도를 분석할 수 있는 N-gram 분석을 실시하였다. Table 2를 통해 다이어트 키워드가 사용될 때 식단 키워드가 동시에 출현한 횟수는 42,914개였다. 두 번째로 다이어트 키워드와 연관이 높은 키워드는 시작 키워드로 37,558개로 나타났다. 세 번째로 다이어트 키워드와 연관이 높은 키워드는 성공 키워드로 35,506개로 나타났다. 단순빈도 분석에서 1위를 한 운동키워드는 N-gram을 이용한 분석에서는 7

Table 1. Frequency of keywords related to diet by simple frequency analysis.

Rank	Keyword	Frequency	Rank	Keyword	Frequency
1	Exercise	178,197	11	Juvis	59,382
2	Health	80,173	12	Breakfast	58,575
3	Menu	79,675	13	Reduction	57,091
4	Effect	79,604	14	Method	54,008
5	Start	76,495	15	Management	46,368
6	Today	70,272	16	Body	46,238
7	Tea	65,882	17	Food	45,670
8	Success	63,236	18	Dinner	45,300
9	Flesh	62,357	19	Diary	40,408
10	Kg	60,705	20	Review	39,932

위로 27,642개 나타났다.

또한 Keyword1과 Keyword2를 이용하여 키워드 간의 방향성을 알아볼 수 있다. Keyword1에서 다이어트 키워드를 제외하고 가장 먼저 건강이라는 키워드가 언급되고 그 다음 keyword2에서 다이어트 키워드가 나타났다. 따라서 건강해지기 위해 다이어트를 실시한다고 해석할 수 있고, 다이어트 행동의 첫 번

째가 식단이라는 것을 방향성을 통해 확인할 수 있었다. 또한 건강을 위해 다이어트를 실시하고, 시작 시기, 성공 방법, 다이어트 효과 등을 궁금해한다고 해석할 수 있다.

Table 2, Frequency of keywords related to diet by N-gram analysis.

Rank	Keyword1	Keyword2	Frequency	Rank	Keyword1	Keyword2	Frequency
1	Diet	Menu	42,914	11	Exercise	Diet	19,034
2	Diet	Start	37,558	12	Weight	Reduction	18,855
3	Diet	Success	35,506	13	Diet	Meal	17,564
4	Diet	Effect	31,274	14	Oriental medicine	Diet	16,899
5	Diet	Tea	28,917	15	Diet	Help	16,207
6	Diet	Diary	28,380	16	Effect	Diet	12,709
7	Diet	Exercise	27,642	17	Diet	Health	12,305
8	Health	Diet	26,448	18	Diet	Stimulus	12,084
9	Diet	Method	24,845	19	Diet	Herbal medicine	12,045
10	Diet	Specialty	19,496	20	Diet	Food	11,949

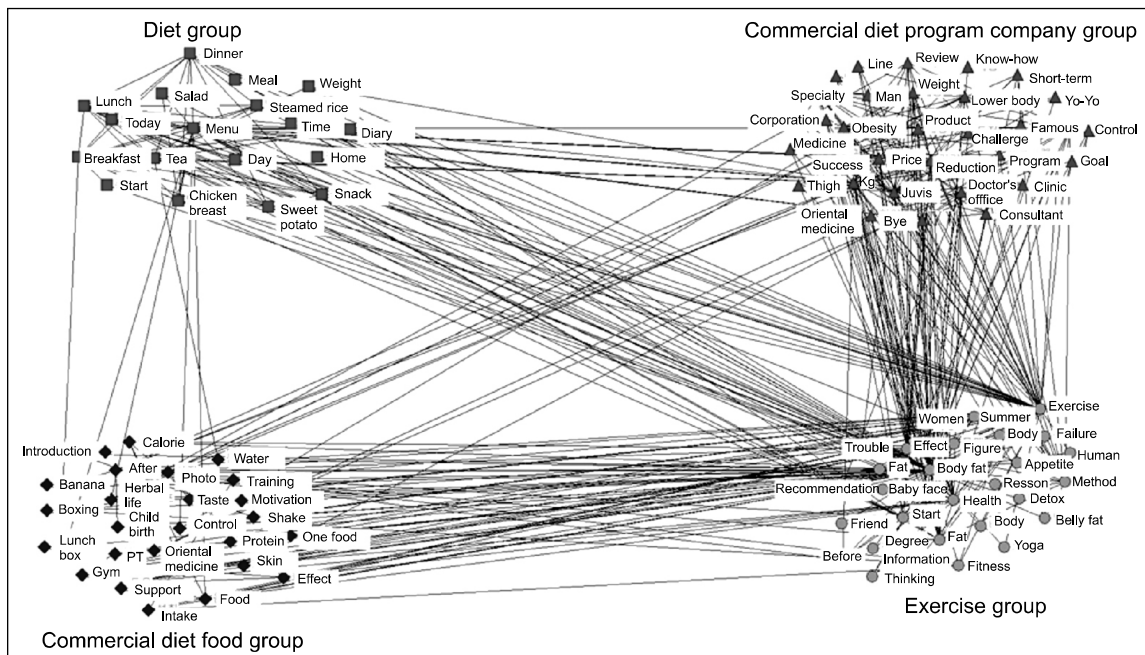


Figure 1. Keyword network analysis related to diet.

### 3. 키워드 네트워크 분석

키워드 네트워크 분석을 위해 본 연구에서는 CONCOR 분석을 이용하였다. Fig. 1은 CONCOR 분석을 나타낸 것이고, 데이터는 4그룹으로 분류되었다. 식이 그룹으로 샐러드, 닭가슴살, 식사, 식단, 아침, 저녁 등의 키워드가 묶여졌고, 운동 그룹으로 몸매, 체지방, 뱃살, 운동, 요가, 헬스 등의 키워드가 묶여졌다. 상업적 다이어트프로그램 그룹으로 기업, 전문, 프로그램, 컨설턴트, 클리닉, 의원 등의 키워드가 묶여졌고, 상업적 다이어트식품 그룹으로 셰이크, 단백질, 식품, 도시락, 원푸드 등의 키워드가 묶여졌다. 네트워크 분석을 통해서 식단과 관련된 키워드와 운동과 관련된 키워드가 분리되어 그룹이 형성된 것을 확인할 수 있다. 또한 상업적 다이어트 그룹도 다른 성격의 2그룹으로 나뉘었고, 상업적 다이어트 프로그램 그룹과 상업적 다이어트식품 그룹으로 형성되었다.

### 4. 계절성 분석

단순빈도 분석을 통해 계절 중에서는 여름이 가장 많이 출현하는 것을 확인하였고 이를 확인하기 위해 다이어트 키워드 빈도를 월별 기간으로 나누어 분석하여 확인하였다. Fig. 2는 다이어트 키워드

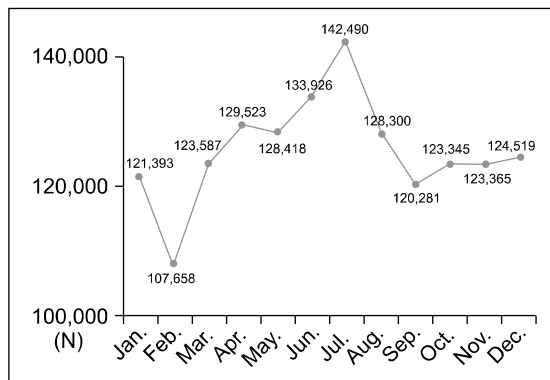


Figure 2. Monthly frequency of keywords related to diet in 2015.

의 월별 출현빈도를 나타낸 그래프로 1월부터 12월 중 다이어트 키워드가 가장 낮게 나온 달은 2월로 107,658개라는 것을 확인할 수 있었다. 그리고 가장 높게 나온 달은 7월로 142,490개라는 것을 확인할 수 있었다. 또한 2월부터 7월까지 다이어트 키워드 빈도가 꾸준히 증가하고, 추석이 끝난 10월이 높아진 것을 확인할 수 있다.

## 고 찰

다수의 연구자들은 필요한 정보를 수집하기 위해서 효과적이고 효율적인 설문조사 방법을 많이 이용한다(Boynton & Greenhalgh 2004). 특히 다이어트에 대해 연구한 기존 논문들은 다이어트 목적, 다이어트 방법 등을 알아보기 위해서 주로 청소년, 여성들을 대상으로 설문조사를 실시하였다(Kwak & Jung 2015). 설문조사는 응답자가 주어진 질문에 맞춰 자신의 생각을 기입하는 방법으로, 이를 통해 응답자에 대한 자료 획득, 문제 파악 및 이해, 예측 등을 알아보고 분석할 수 있다(Alreck & Settle 2004; Jeon & Choi 2005). 그러나 설문조사는 많은 비용과 시간, 조사자 등이 필요하고, 설문지에 제시된 질문에만 답변을 해서 답변이 다양하지 않으며 폐쇄적이고, 종종 설문 응답자들은 본인이 참여한 설문조사에 흥미나 관심이 없으며, 설문지 문항을 제대로 이해하지 못하는 경우도 있기 때문에 결과를 신뢰하기 어렵다는 문제점도 있다. 또한 모집단의 정의, 표본 추출 방법, 표본 크기 등 표본 추출을 위해 표집설계를 수행해야 하는데, 이를 잘못 반영하면 추출된 표본이 모집단 전체를 대표한다고 간주할 수 없고, 편향된 신뢰구간, 표준오차 등 한계점으로 인해 잘못된 결과를 얻을 수 있다(Seol & Chung 2000). 따라서 본 연구는 이를 해결하고자 빅데이터를 통한 새로운 분석 방법을 시도하고자 하였다. 우선 빅데이터를 이용하면 인터넷을 활용하는 사람들이 집단이 되기 때문에 설문조사 시 추출된 집단의 범위와

달라 전수조사 개념에 더 근접하게 된다. 또한 자신의 생각이나 느낌을 자발적으로 인터넷에 작성한 자료들을 활용하기 때문에 설문조사와 달리 다양한 결과를 도출할 수 있다(Kang 2013). 이에 본 연구에서는 네이버에 있는 블로그, 카페, 웹문서, 뉴스에서 다이어트 키워드가 포함된 문장을 토대로 단순빈도 분석, N-gram 분석, 키워드 네트워크 분석 그리고 계절성 분석을 진행하였다.

단순빈도 분석에서는 ‘운동’ 키워드(178,197개)가 1위로 나타났고, 운동 키워드가 해당하는 문장들을 찾아 살펴보니 ‘다이어트 운동으로는 복근운동, 다리 운동, 순환운동 등 다양한...’ 등 한 문장에 운동이라는 단어가 여러 번 반복되어 출현빈도가 높게 나왔음을 확인할 수 있었다. 또한 아침(58,575개), 점심(37,549개), 저녁(45,300개) 등 식단과 관련된 키워드의 출현빈도가 높았으며, 특히 사람들이 아침에 대해 높은 관심을 보이고 있음을 알 수 있다. Lee(2003)는 아침식사 유형에 대해 설문조사하였으며, 본 연구 결과에서도 점심이나 저녁식사보다 아침식사에 관한 출현빈도가 높게 나왔다.

단순빈도 분석은 전체 데이터에서 단순 출현빈도만을 나열하기 때문에 키워드와 키워드 간에 얼마나 밀접되어 있는지 분석할 수 없다. 따라서 포털사이트의 검색어 자동 완성과 유사한 개념인 N-gram 분석을 통해 결과를 확인해보니 다이어트-식단(42,914개)이 가장 높게 나타났고, 이는 전체 데이터에서 다이어트 키워드가 사용될 때 식단 키워드가 동시에 연쇄적으로 출현한 횟수가 가장 많은 것을 의미한다. 따라서 다이어트 키워드와 가장 관련이 밀접한 단어는 식단이라고 볼 수 있다. 그리고 다이어트-시작(37,558개), 다이어트-성공(35,506개), 다이어트-효과(31,274개) 등의 순으로 나타났다. 또한 본 연구에서는 다이어트-식단(42,914개), 다이어트-차(28,917개), 다이어트-식품(17,564개), 다이어트-음식(11,949개) 등이 분석되어 다이어트는 식이와 관련되어 있으며, 다이어트 시 사람들은 식이에 대해 관심이 높다고 사료된다. N-gram 분석의 키워드 위치에 따라 방향성

이 결정되어 Keyword1과 Keyword2의 방향성을 통해 건강 및 비만에서 다이어트로 향해가는 결과를 얻었고, Kim 등(2015)의 대학생 대상의 설문조사에서도 다이어트를 목적으로 날씬한 몸매(1위), 체중조절(2위), 건강(3위) 등을 발견하였다. 즉, 사람들이 다이어트를 하는 목적은 건강을 위해서 또는 비만을 벗어나기 위해서 다이어트를 시도한다고 간주할 수 있다. 하지만 본 연구의 결과에서는 체중감량보다 건강, 다이어트 키워드의 출현빈도가 높게 나왔으며 이는 조사 대상자가 대학생으로 한정되어 있지 않고 인터넷을 사용하는 사람 전체이기 때문에 사료된다. 이에 다이어트 홍보 시 체중조절이나 몸매를 강조하는 홍보도 좋지만 많은 사람들에게 효과적으로 홍보하기 위해서는 건강을 초점으로 홍보해도 좋을 것으로 사료된다.

또한 다이어트를 위한 행동으로 사람들은 제일 먼저 식단을 고려한다고 간주할 수 있다. 하지만 다이어트에 어떤 식단이 도움되는지에만 초점을 두고 있고, 섭취하는 양에 대해서는 관심을 가지지 않는다고 간주할 수 있다. Stroebele & De Castro(2004)는 음식을 선택할 때 무엇을 섭취할까보다 얼마만큼 섭취할지 섭취량을 결정하는 것이 중요하다고 하였다. 따라서 다이어트를 시도하는 사람들이 식단에 대한 관심을 가지는 것도 중요하지만 섭취해야 하는 양도 동시에 인지할 수 있는 방향으로 다이어트 방법이 개선되어야 할 것이다.

또 다른 다이어트 행동으로 자신이 스스로 일기를 작성하여, 이를 통해 언제, 어디서, 무엇을 얼마만큼 먹었는지 작성한 후 기억할 수 있다. 다이어트 일기를 작성하면 자신의 식행동을 객관적으로 관찰할 수 있기 때문에 스스로 모니터링을 할 수 있게 되고 다이어트에 도움이 될 것으로 사료된다. Han 등(2004)은 비만 치료 시 식사일기를 작성하도록 하였으며, 작성 여부에 따라 식사일기를 작성한 군과 작성하지 않은 군으로 나누어 비만 치료변화량을 측정해보니 유의하게 차이가 나타났다고 보고했다. 따라서 다이어트 시 일기를 작성하는 것이 도움이

되지만 Han 등(2004)의 연구처럼 식사일기를 꾸준히 작성하지 못하는 사람들이 존재한다. 이는 일기를 작성하기가 번거롭고 귀찮기때문으로 사료된다. 이를 해결하기 위해 Lee & Chang(2016)은 스마트폰의 앱을 이용하여 효율적으로 기록할 수 있는 식사일지를 제안하였다. N-gram 분석을 통해 다이어트와 관련된 키워드의 1차원적인 결과와 결과의 연장선으로 다이어트의 개선점과 새로운 접근법을 파악하는데 도움이 될 것으로 사료된다.

N-gram 분석을 통해 다이어트 키워드와 관련된 키워드를 알아보고, 키워드 네트워크 분석을 통해서 다이어트 키워드와 관련된 키워드의 네트워크 관계를 확인하고자 했다. 키워드 네트워크 분석을 통해 키워드 간의 그룹은 총 4그룹으로 분류되었고, 그룹의 특징을 보면 식이 그룹, 운동 그룹, 상업적 다이어트프로그램 그룹, 상업적 다이어트식품 그룹으로 형성되었다. 이는 다이어트 방법들이 군집화된 것으로 판단된다. Jung 등(2015)의 여고생 대상의 설문조사와 Choi(2015)의 여대생 대상의 설문조사 결과에 의하면 다이어트 방법으로 식이 조절, 운동 조절이 대다수를 차지했고 상업적 다이어트 방법은 소수에 불과했다. 본 연구의 그룹 또한 식이 그룹과 운동 그룹이 나타났지만 상업적 다이어트프로그램 그룹도 나타났다. 이는 폐쇄형 답안이 아닌 개방형 자료를 대상으로 분석하여 다양한 그룹이 형성된 것으로 사료되며, 대상이 여고생, 여대생이 아닌 학생, 직장인, 주부 등을 포함한 다양한 연령대 그룹으로 형성되어 경제력 차이가 있기 때문으로 생각된다. Won 등(2012)은 상업용 다이어트식품을 구매하는 만 25~45세 성인을 대상으로 설문조사를 한 결과 대상의 88.1%가 월평균 소득이 300만원 이상이었다고 보고했다. 따라서 경제력이 낮은 여학생 대상의 설문조사에서는 상업적 다이어트 결과가 적게 나왔지만, 본 연구는 다양한 연령대를 대상으로 했기 때문에 상업적 다이어트 결과가 많이 나온 것으로 사료된다.

마지막으로, 다이어트는 계절에 민감할 것으로 판

단되어 다이어트 키워드 출현빈도를 통한 계절성 분석을 실시하였는데 1년간 1,506,805개가 출현하였고 월평균은 125,567개 나왔다. 이를 통해 2월을 제외하고는 사람들이 꾸준히 다이어트에 관해 관심을 가진다고 생각된다. Pack(2004)은 날씨가 따뜻해지면서 옷이 얇아지고 노출의 계절이 시작되면 사람들의 체중감량 시도가 활발해진다고 발표하였고, 본 연구의 결과에서도 겨울이 끝나는 2월부터 7월까지 다이어트 출현빈도가 꾸준히 증가하고 있는 것을 확인하였다. 그리고 Lee(2015)는 G마켓에서 발표한 월별 다이어트용품 구매 시기를 토대로 다이어트용품 구매빈도가 2월부터 7월까지 꾸준히 증가한다고 발표하였고, Kim(2015)은 이마트 정보를 바탕으로 3월부터 다이어트 관련 상품의 구매빈도가 상승하였고 5월에는 가파른 상승을 하였다고 보고하였다. 이와 마찬가지로 본 연구의 결과도 2월부터 7월까지 다이어트 출현빈도가 높아지는 것을 확인할 수 있었으며, 8월부터 점차 줄어들다가 10월에 다시 증가하였다. 이는 2015년의 9월 26일부터 28일까지 추석연휴였고, 추석이 끝난 10월부터 다이어트를 시작하려는 사람이 많아진 것으로 사료된다.

## 요약 및 결론

본 연구에서는 우리나라 사람들의 다이어트에 대한 경향 및 인식을 알아보기 위해서 설문조사 대신 빅데이터 분석 방법을 이용하였고, 포털 사이트 네이버를 통해 2015년 1월 1일부터 2015년 12월 31일까지 1년간 다이어트 키워드가 포함된 문장을 분석하여 단순빈도 분석, N-gram 분석, 키워드 네트워크 분석, 계절성 분석을 시행하였으며 결과는 다음과 같다.

1. 단순빈도 분석을 통해 다이어트 키워드가 포함된 문장 중 다이어트 키워드를 제외하고 가장 많이 출현한 키워드는 ‘운동(178,197개)’으로 나타났지만, 키워드 간의 연관빈도를 분석한 N-gram 분석에서는 다이어트-식단(42,914개)이 가장 많이 출



현했다. 따라서 다이어트 시 운동보다는 다이어트 식단이 중요하게 나타났다.

2. N-gram 분석의 키워드 방향성을 통해 사람들은 건강을 생각하여 다이어트를 시작하고자 하고, 다이어트 행동으로 제일 먼저 식단에 대해 고민한다고 나타났다.
3. 다이어트 키워드와 연관된 키워드를 유사한 성격 들끼리 그룹화한 키워드 네트워크 분석을 통해 총 4개의 그룹으로 분류되었고, 식이 그룹, 운동 그룹, 상업적 다이어트프로그램 그룹, 상업적 다이어트식품 그룹이었다.
4. 계절성 분석을 통해 사람들은 노출이 심해지는 7월에 다이어트에 대해 관심이 가장 많았고, 명절이 끝난 다음 달에 관심이 증가하는 것을 확인하였다.

이상의 결과에서 사람들은 건강 때문에 다이어트를 시작하고, 다이어트 실천을 위해서 우선적으로 식단에 대해 고민하며, 주로 여름에 다이어트에 관한 관심이 가장 높은 것으로 나타났다. 또한 다이어트는 유행에 민감하기 때문에 다이어트 유행을 파악하기 위해서 연도별 빅데이터 분석을 시행하면 포괄적으로 전체를 알아볼 수 있고 비교할 수 있을 것으로 기대된다.

## REFERENCES

- 마경근 (2015): 빅데이터를 활용한 과학적 행정구현. 지역정보화지 91:14-19
- Ahn HJ (2012): Extraction of keywords from brand images texts using network analysis. JKIIIT 10(2):176-182
- Ahn MS, Oh IK (2015): Analysis of attitudes on using five-star hotel packages applying network text analysis method. JKASTM 30(5):163-181
- Alreck PL, Settle RB (2004): The survey research handbook. 3rd ed. McGraw-hill/Irwin press. Boston. pp.3-55
- Boynton PM, Greenhalgh TA (2004): Selecting, designing, and developing your questionnaire. BMJ 328(7451):1312-1315
- Chang U, Ha J, Hong W (2004): The trial status and expectation degree on internet diet program by female university students : focused on trial status. J Korean Diet Assoc 10(3):356-363
- Choi HJ (2015): A comparative study of Korean and Japanese female college students' body image, dieting methods and exercise habits: focusing on comparison among local female college students. J Korean Soc Living Environ Sys 22(6): 927-939
- Diesner J, Carley KM (2004): Causal mapping for research information technology. Idea Group Publishing. London. pp.81-83
- Han JY, Cho JH, Jang JB, Lee KS (2004): The effects of the diary for diet and exercise in hospital with lipodren and auricular acupuncture therapy on the treatment of obesity. J Korean Med Obes Res 4(1):213-219
- Herman CP, Polivy J (1988): Psychological factors in the control of appetite. Curr Concepts Nutr 16:41-51
- Hwang MG, Choi DJ, Lee HG, Choi C, Ko BK, Kim PK (2012): Domain N-gram construction and its application. JKIISE 37(2):47-51
- Jeon CH, Choi HK (2005): Implementation of respondent-based real-time survey system using XML. JKIISE 30(1):648-650
- Jung A, Ryu H, Song K, Lee H (2015): A comparison of dietary habits, weight control behaviors, eating disorder risk, and depression of middle school girls according to various stages of dieting. Korean J Community Nutr 20(3):178-187
- Kang SJ (2013): A study on direction for application of big data in design. JKSDC 19(3):1-12
- Kim JW (2015): Emart, when early diet products special. Available from: <http://www.ewnewsday.co.kr/news/articleView.html?idxno=382183>. Accessed September 16, 2016
- Kim MH, Kim YJ, Chung JS, Yeon JY (2015): Fad diet status of male and female collegians. Korean J Food Nutr 28(2): 258-268
- Koh BK, Yoon JS (2003): Current tendency of middle school students of get the food and nutrition information from the internet web site. J Korean Soc Food Sci Nutr 32(1):102-108
- Kwak SH, Jung EC (2015): Understanding diet behavior : focusing on the expanded theory of planned behaviors. J Commun Sci 15(4):5-56
- Lee JS (2003): A study of female college students' behavior and ideal breakfast types. Korean J Food Culture 18(5):466-474
- Lee JY, Chang UJ (2016): Development scenario of dietary intake survey using internet of things (IoT). J Korean Diet Assoc 22(3):225-231
- Lee OJ, Park SB, Chung DU, You ES (2014): Movie box-office

- analysis using social big data. *J Contests Assoc* 14(10):1-12
- Lee YM (2015): Year-end is free diet...snacks for sale ↑ ·diet products ↓ . Available from: <http://www.yonhapnews.co.kr/bulletin/2015/12/07/0200000000AKR20151207179900030.HTML?input=1195m>. Accessed September 16, 2016
- Moon JY (2013): A study of the intellectual structure of secretarial studies using network analysis. *J Secr Sci* 22(1):125-145
- Pack HJ (2004): Various diet methods currently in vogue among the individual. *Korean J Obes* 8:470-471
- Seol HS, Chung TH (2000): A study on methods estimating standard errors using replicated sampling. *J Korean Educ* 27(1):98-111
- Shin MY (1997): A process that mass-media affects dieting and eating disorder of young women. Masters degree thesis. Kangwon National University. pp.4-5
- Stroebele N, De Castro JM (2004): Effect of ambience on food intake and food choice. *Nutrition* 20(9):821-838
- Turner S (1996): A healthy weight-loss program includes provider involvement and client acceptance. *Nurse Pract* 21(4): 161-162
- Van JO (2007): Food attitude and food behavior of female consumers according to their lifestyle. Doctors degree thesis. Kyonggi University. pp.61-64
- Wasserman S, Faust K (1994): *Social network analysis : methods and applications*. Cambridge University Press. Cambridge. pp.370-374
- Won HS, Lee HJ, Kwak JS, Kim JH, Kim MK, Kwon OR (2012): Study on purchase and intake patterns of individuals consuming dietary formula for weight control or health/functional foods. *Korean J Nutr* 45(6):541-551
- World Economic Forum (2012): The top 10 emerging technologies for 2012. Available form: <https://www.weforum.org/agenda/2012/02/the-2012-top-10-emerging-technologies/>. Accessed September 1, 2016