

원저

건강정보 소비자 용어 시스템 구축을 위한 기반연구

신윤주¹, 김정은^{1,2}

서울대학교 간호대학¹, 서울대학교 간호과학연구소²

A Study for Building a System of Consumer Vocabulary for Health Information

Yoonju Shin¹, Jeongeun Kim^{1,2}

College of Nursing, Seoul National Univ.¹,
Research Institute of Nursing Science, Seoul National Univ.²

Abstract

Objectives: The purposes of this study were to identify the difference between consumer vocabulary and medical vocabulary in terms of health information; to understand the features of consumer vocabulary; and to contribute by building a system that is able to link consumer vocabulary with medical vocabulary. **Methods:** Data collection was conducted using articles in the knowledge corner of a portal web-site. A total of 43,304 health-related terms (total terms extracted) were collected as objects of this study and these terms were analyzed for their mapping rate and frequency of use (the repeated number of a term). **Results:** The rate of mapping between the consumer vocabulary for health information and the medical vocabulary was not high. However, the number of “unmapped terms” was decreased by linking terms having similar forms to “preferred terms” and by extending synonyms. **Conclusion:** Linking with preferred terms and extending synonyms are, thus, required to increase the mapping rate between consumer vocabulary for health information and medical vocabulary, and the terms that consumers use are essential to further be researched in order to understand their morphology and features; hence, increasing consumer accessibility to the medical field. (*Journal of Korean Society of Medical Informatics 15-1, 31-40, 2009*)

Key words: Consumer Vocabulary, Medical Vocabulary, Health Information, Mapping, Preferred Terms, Synonyms

논문투고일: 2008년 10월 9일, 심사완료일: 2009년 2월 11일

교신저자: 김정은, 서울시 종로구 연건동 28번지 서울대학교 간호대학 508호 (110-799)

Tel: 02-740-8483, Fax: 02-765-4103, E-mail: kim0424@snu.ac.kr

I. 서론

과거 대부분의 건강정보와 관련된 연구와 개발은 의료전문가 중심으로 이루어졌으며 병원 중심 활동에 초점이 맞추어졌다. 그러나 최근 소비자들이 질적인 삶을 유지하는데 도움이 되는 유용한 정보를 찾기 위해 적극적으로 행동함에 따라 건강정보 관련 연구는 소비자의 요구를 포함하는 방향으로 점차 변화하고 있다. 이로 인하여 소비자들의 건강에 대한 관심과 요구에 부응하는 방대하고 다양한 건강정보가 제공되고 있으며 건강정보 소비자와 의료전문가들을 위한 많은 인터넷 웹사이트와 응용프로그램이 개발되고 있다¹⁾.

소비자는 유용한 건강정보의 습득을 통해 질병을 이해하고 자가관리 능력을 향상시키며 건강과 관련된 의사결정 과정에도 참여할 수 있으나 이러한 건강정보로의 접근은 익숙하지 않은 의료전문가 용어와 의료분야의 지식부족으로 많은 어려움이 있다²⁾. 건강정보 소비자들이 주로 이용하는 정보는 의료전문가들이 사용하는 용어 중심인 반면에, 건강정보 소비자들은 자신의 의사를 표현하기 위해 의료전문가 용어와 일상용어를 조합하여 사용하며 사회심리적, 문화적 편차에 의해서 의료전문가 용어와 더욱 큰 차이를 보이기도 한다³⁾. 또한 이러한 표현의 다양성은 “용어 문제”로 알려진 “잘못된 의사소통”을 야기할 수 있다⁴⁾.

특히 중간 매개체 없이 웹을 통하여 건강정보에 접근하는 소비자들은 스스로 이해하고 지식화하는 단계에서 의료정보 관련 개념과 용어를 직접 접하게 된다. 따라서 건강정보 소비자는 익숙하지 않은 의료전문가 용어를 읽고 이해하는 과정에서 소비자 용어와 의료전문가 용어의 차이로 인해 정보 제공자가 의도하지 않은 내용으로 용어를 잘못 연상하거나 이해할 수 있으며, 비록 건강정보 소비자가 의료전문가와 동일한 용어를 사용할지라도 의료분야에서의 개념보다는 소비자들이 실제 사용하는 개념과 연관 지을 것이다. 또한 건강정보 소비자의 질의어가 데이터베이스 내의 색인(Index)과 다른 경우 정보가 검색되지 않으며, 검색하고자 하는 정보에 대한 정확한 지식이 없다면 대부분의 소비자는 자신의 정보 요구와 정확하게 일치하는 질의어를 작성하는데 어려움이 있게 된다⁵⁾.

이처럼 건강정보 소비자와 의료전문가의 건강관련 용어의 차이는 상호간의 효과적인 의사소통과 건강정보 검색, 그리고 정보화된 의사결정을 가로막고 있다

³⁾ 또한 용어의 차이는 건강정보 소비자의 검색과 자가관리에 대한 책임 증가에 따라 더욱 심각한 문제가 되고 있으며 나아가 이해하기 어려운 의료전문가 용어는 소비자의 건강정보 이해능력에 장애가 되어 기대하는 건강결과를 얻기 어렵게 된다.

이에 본 연구는 건강정보 소비자 용어 시스템 구축을 위한 기반연구로서 먼저, 건강정보 소비자 용어와 의료전문가 용어와의 차이를 규명하고 건강정보 소비자 용어의 특징을 분석하고자 시도되었다.

II. 재료 및 방법

1. 연구설계

본 연구는 인터넷 게시물에서 건강정보 소비자 용어를 추출 후 의료전문가 용어와의 매핑을 통하여 두 용어의 차이 및 소비자 용어의 특징을 파악함으로써 건강정보 소비자 용어와 의료전문가 용어를 연결시킬 수 있는 시스템 개발을 위한 기초자료를 생성하기 위한 조사 연구이다.

2. 연구대상

게시물 수, 접근성, 대표성, 현재 이용도를 기준으로 인터넷 포털사이트(www.yahoo.co.kr) 지식코너의 건강 Q&A 게시물에서 건강과 관련된 개념(의미)을 나타내는 모든 용어를 추출 대상으로 하였으며 일반 소비자의 관점을 반영하기 위하여 비의료인으로 구성된 4명의 용어 추출자에 의해 추출된 43,304개의 용어를 최종 연구대상으로 하였다.

3. 연구방법

자료 수집은 정확도와 인기도, 소비자 게시물 수, 접근성, 대표성, 이용도를 기준으로 하여 의도표출(purposive sampling) 하였다.

(1) 소비자 용어와 의료전문가 용어(통제어휘) 대상 선정 및 추출

소비자 용어 대상으로 인터넷 포털사이트(www.naver.com)에서 ‘건강정보, 의료정보’ 두 검색어를 기준으로

정확도와 인기도에 따라 300개의 건강정보제공 사이트를 사전조사 후 2차로 소비자 게시물 수를 기준으로 27개 사이트-인터넷 포털사이트 지식코너(네이버 지식인과 야후 지식), 국민건강보험공단 건강인, 보건사회연구원 건강길라잡이, 3차 종합전문병원 홈페이지(20개 병원), 건강샘, 다음과 네이버 건강·의학 카페-를 선정하였다. 3차로 게시물 수, 접근성, 대표성, 현재 이용도를 기준으로 인터넷 포털사이트 지식코너 (<http://kr.ks.yahoo.com>)에 올려진 게시물을 최종 자료 수집 대상으로 선정하였다. 통제어휘 대상으로는 의료전문가 집단을 대표하는 대표성, 접근성, 용어 수를 기준으로 우리나라 의학계를 대표하는 기구인 대한의사협회에서 발간한 의학용어집을 최종 선정하였다.

2007년 1월 1일부터 2007년 6월 30일 사이에 일반 소비자(비의료인)에 의해 올려진 건강정보 게시물에서 건강(질병 포함)과 관련된 개념을 나타내는 말뭉치(corpus)를 4명의 일반인(비의료인)이 추출하였다. 용어 추출과정에서 동일내용으로 올려진 게시물, 의료인에 의해 작성된 게시물, 한의학 및 민간요법 주제

의 게시물 및 의료인에 의해 작성되는 경우가 많은 답글은 제외되었으며 학회 및 공공기관 외의 고유명은 생략하였다. 또한 개념을 나타내는 최소 단위의 말뭉치에서 의미에 영향을 미치지 않는 조사는 편의상 추출단계에서 생략하였으며 본 연구의 통제어휘 형태와 같이 명사형 어미로 정리하였다.

(2) 소비자 용어의 분류 및 우선어 선정

본 연구에서 사용된 용어 중 표준용어사전에 없는 일부 용어는 국립국어원 표준국어대사전과 한국정보통신기술협회 정보통신용어사전 표준화 담당자, 의료정보 전문가의 자문을 받아 Table 1과 같이 정의하였다.

7,036개 게시물에서 추출한 43,304개 용어 중 중복되는 동일한 용어들을 자체 개발한 전산프로그램을 이용하여 용어명과 사용빈도(예로 추출한 용어 중 “병원” 용어의 200회 중복시 추출 용어는 “병원”, 추출 용어의 사용빈도는 “200회”로 정리함)로 각각 정리하여 4,468개의 추출 용어를 분석의 대상으로 하였다.

Table 1. Definition of terms

Terms	Definition
표본(Samples)	전체 추출된 용어들
사용빈도(Use frequency)	용어의 반복된 횟수
추출 용어(Extracted terms)	중복된 동일한 용어를 제외한 추출된 용어
추출 용어의 사용빈도(Use frequency of extracted terms)	추출 용어의 반복된 횟수
우선어(Preferred terms)	동일개념을 갖는 용어들을 대표하는 용어
우선어의 사용빈도(Use frequency of preferred terms)	우선어가 대표하는 용어들의 반복된 횟수 합
일치용어(Matched terms)	추출 용어 중 우선어와 일치하는 용어
불일치 용어(Unmatched terms)	추출 용어 중 우선어와 일치하지 않는 용어

Table 2. Criteria for selection of Preferred terms

Category	Extracted terms	Criteria	Preferred terms	Matched terms	Unmatched terms
Terms of similar morphology	가느다란 변 가는 변 가는 대변	국어사전 추출빈도	가는 변	가는 변	가느다란 변 가는 대변
Word of foreign origin	임플란트 임플란트 임플런트	국어사전 추출빈도	임플란트	임플란트	임플란트 임플런트
Obscure concept	정액 하얀 분비물 투명한 액체	국어사전	정액	정액	하얀 분비물 투명한 액체
Subordinate concept Included figure	임신 8주 임신 18주		임신		임신 8주 임신 18주
Shortened word	자궁내막증 내막증	국어사전	자궁내막증	자궁내막증	내막증

매핑률을 높이기 위한 작업으로써 추출 용어를 유사형태 용어, 외래어, 동일개념 용어, 숫자가 포함된 하위개념, 줄임말의 5개 영역으로 구분하여 국어사전(국립국어원 표준국어대사전), 의학사전(대한의사협회 의학용어집 외 본 연구 통제어휘 대상)과 추출 빈도를 기준으로 동일 개념의 용어들을 대표할 우선어를 선정하였다(Table 2). 추출 용어 중 우선어와 동일한 용어들을 일치 용어로 구분하여 정리하였으며 동의어 및 상하위 개념의 용어들은 우선어 선정 대상에서 제외하였다(Fig. 1).

(3) 소비자 용어와 의료전문가 용어 매핑

1차로 전산프로그램을 이용하여 추출 용어와 우선어를 통제어휘 대상인 대한의사협회 의학용어 4집과 매핑을 실시하였다. 매핑 과정에서 띄어쓰기는 인식되지 않도록 프로그램을 구성하였고, 의학용어집의 ‘방광절제(술)’ 형태의 용어들은 ‘방광절제’와 ‘방광절제술’로 각각 매핑되었다. 대한의사협회 의학용어 4집에서는 의학용어의 한글화 과정이 이루어졌는데, 이로 인해 소비자들이 많이 사용하는 옛 표기방식이 용어집에서는 제외되어 통제어휘로써 용어 수의 부족함을 보완하기 위해 2차 매핑에서는 1차 매핑에서 매핑되지 않은 용어를 대상으로 KMLE(www.kmle.co.kr)

에서 대한의사협회 필수용어집, 옛 대한의사협회 의학용어사전, 대한해부학회 의학용어사전, 대한신경외과학회사전, 대한기생충학회 의학용어사전, 대한생화학분자생물학회 용어사전, 한국표준질병사인분류약자, 경북대 치과대학 구강내과교실사전 검색을 실시하였다. KMLE 웹사이트에서 검색되지 않은 약품 용어는 대한약사회(www.kpanet.or.kr)의 약품 검색을 통해 추가 매핑하였다.

또한 매핑되지 않은 소비자 용어와 의료전문가 용어의 연결을 위하여 우선어의 동의어 확장을 시도하였다. 국어사전(국립국어원 표준국어대사전), 백과사전(두산백과사전), 의학용어집(대한의사협회 의학용어집 외 통제어휘 대상 사전) 내의 동의어 관계를 참조하여 동의어 선정 후 의료전문가 2인으로부터 감수를 받았다. 우선어의 상당수가 개념의 범위가 넓은 용어들(예로 안과 수술, 갑상선 증상)이었기 때문에 동의어 확장은 제한적으로 이루어졌다(Fig. 2).

(4) 소비자 용어의 의미적 분석

미국의 UMLS 시맨틱 그룹⁴⁾을 참고하여 우선어를 신체(유전/호르몬/효소/신경/세포/미세구조 제외), 질병명, 증상, 치료, 기관/조직, 화학/약, 행동, 검사, 기계, 물품, 유전/호르몬/효소/신경/세포/미세구조, 직업,

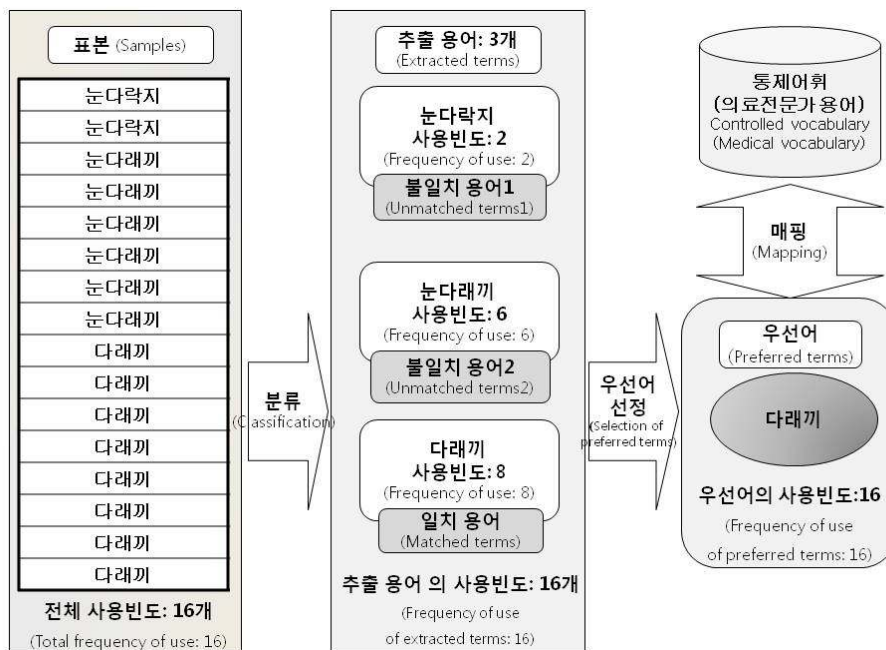


Figure 1. Steps for classification of consumer vocabulary and selection of preferred terms

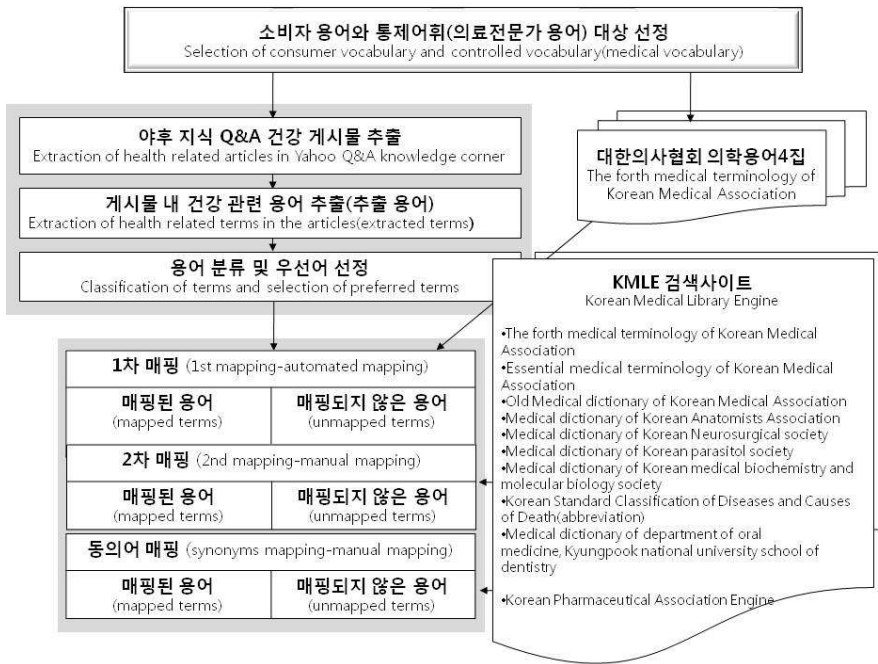


Figure 2. Mapping process between consumer vocabulary and medical vocabulary

식이/식품, 인간, (미)생물/세균, 기타 개념 16개 범주로 분류 후 소비자 건강정보 전문가로부터 감수를 받았으며, 범주별 용어와 통제어휘와의 매핑 결과로 공통된 특징 및 용어의 전반적인 분석을 실시하였다.

4. 자료분석방법

수집된 자료는 자동화된 매핑 전산프로그램, KMLE와 대한약사회 웹사이트 용어검색, 사전 및 용어집 내 동의어 관계 수작업, MS Office 2003 Excel Program을 이용하여 실수 및 백분율을 구하고, 범주별로 분류하여 분석하였다.

Ⅲ. 결과

1. 소비자 용어의 전반적 분석결과

추출 용어 4,468개, 우선어 3,819개, 추출 용어와 우선어가 일치한 일치 용어 3,563개, 일치 용어의 사용빈도 39,982회, 추출 용어와 우선어가 일치하지 않은 불일치 용어 905개, 불일치 용어의 사용빈도 3,322회로 나타났다. 추출 용어의 평균 글자수는 3.90개이고 우선어의 평균 글자수는 3.85개로 우선어 선정과정을

거치면서 글자수가 감소하였다.

추출 빈도는 1회 추출된 우선어가 42.3%(1,614개)였고 2회 추출된 우선어가 16.2%(617개)로서 전체 우선어의 58.5%를 차지하여 많은 소비자 용어에서 낮은 사용빈도를 보였다.

2. 소비자 용어와 의료전문가 용어 매핑

소비자 용어(우선어)와 의료전문가 용어(통제어휘)와의 1차 매핑에서 우선어의 30.7%(1,171개), 2차 매핑까지(1차와 2차 매핑결과 합) 41.2%(1,574개), 동의어 매핑까지(1차, 2차, 동의어 매핑결과 합) 46.3%(1,768회)의 매핑률 증가를 보였으며, 매핑된 우선어의 사용빈도 또한 1차 매핑 64.3%(27,826회), 2차 매핑까지 70.4%(30,493회), 동의어 매핑까지 80.0%(34,636회)로 통제어휘의 증가와 동의어 확장시 매핑률의 증가를 나타냈다(Table 3).

2차 매핑까지의 우선어 매핑률(41.2%)은 추출 용어 매핑률(34.1%)과 비교하여 7.1% 높게, 매핑된 우선어의 사용빈도(70.4%)는 추출 용어의 사용빈도(67.1%)와 비교하여 3.3% 높게 나타났다(Table 4).

매핑된 소비자 용어 중 사용빈도가 높았던 우선어는 다이어트(1,310회), 병원(1,277회), 임신(1,032회),

가슴(866회), 수술(774회), 키(725회), 치료(606회), 여드름(541회), 아픔(459회), 약(427회) 등의 순이었으며, 매핑되지 않은 용어 중 사용빈도가 높았던 우선어는 생리(1,223회), 몸무게(472회), 살빼기(364회), 뱃살(335회), 복용(227회), 냉(172회), 라식수술(150회), 살찜(138회), 아랫배(126회), 산부인과(110회) 등이었다.

3. 소비자 용어의 의미적 분석 및 특징

(1) 소비자 용어의 의미상 범주 분류

의미상 범주별 용어 수는 질병명(735개), 증상(708

개), 신체(556개), 치료(467개) 순으로 높게 나타났으며 용어의 매핑률은 질병명(61.2%)과 신체(55.2%) 범주에 비해 증상(24.0%), 치료(17.1%), 검사(23.2%) 범주 용어의 매핑률이 낮게 나타났다. 유전/호르몬/효소/신경/세포/미세구조 범주의 용어 수(85개)와 사용빈도(0.9%)는 낮았으나 용어의 매핑률(70.6%)은 높게 나타났다(Table 5).

(2) 소비자 용어의 의미적 패턴

전문용어는 용어의 모호함을 피하기 위하여 이상적으로 하나의 개념에 하나의 형태나 용어로 나타내

Table 3. The result of mapping consumer vocabulary and medical vocabulary

	Mapped Preferred terms N=3,819(number)	%	Use frequency of mapped preferred terms N=43,304(time)	%
1st mapping	1,171	30.7	27,826	64.3
1st+2nd mapping	1,574	41.2	30,493	70.4
1st+2nd+synonyms mapping	1,768	46.3	34,636	80.0

Table 4. Comparison with mapping rate of extracted terms and preferred terms

	Extracted terms N=4,468(number) Preferred terms N=3,819(number)	%	Frequency of use N=43,304(time)	%
1st mapped extracted terms	1,181	26.4	26,514	61.2
1st mapped preferred terms	1,171	30.7	27,826	64.3
1st+2nd extracted terms that mapped	1,522	34.1	29,059	67.1
1st+2nd preferred terms that mapped	1,574	41.2	30,493	70.4

Table 5. Mapping rate of terms by classification of meaning

	Terms (number)	Mapped terms (number)	Mapping rate of terms(%)	Use frequency of terms(time)	Use frequency of mapped terms (time)	Mapping rate of use frequency (%)
Disease or Syndrome	735	450	61.2	5,915	5,148	87.0
Symptoms	708	170	24.0	5,610	3,110	55.4
Anatomy(excluding genes etc)	556	307	55.2	10,814	8,507	78.7
Cure	467	80	17.1	3,862	2,266	58.7
Physiology	255	124	48.6	4,793	2,709	56.5
Chemicals/drugs	249	84	33.7	1,689	1,221	72.3
Examination	220	51	23.2	1,551	773	49.8
General concepts/Ideas	197	104	52.8	2,236	1,935	86.5
Devices/goods	94	36	38.3	759	518	68.2
Genes/hormone etc	85	60	70.6	384	324	84.4
Activities&Behaviors	72	30	41.7	2,239	1,184	52.9
Organizations/group	60	26	43.3	2,431	2,040	83.9
Human	38	16	42.1	288	199	69.1
Diet/food	34	13	38.2	299	184	61.5
Living Beings/germs	25	13	52.0	82	53	64.6
Occupations	24	10	41.7	352	322	91.5
Total	3,819	1,574		43,304	30,493	

는 반면에 소비자 용어는 일상언어로 문맥에 의존하여 상이한 개념을 동일한 용어로 표현하기도 하였다. 또한 질과 양 개념의 용어 사용시 주관적인 단위 사용률이 높았고, 용어를 정확히 알지 못하는 경우 은유 및 환유, 축약, 서술적 방법으로 대상을 표현하거나 개념을 정의하였으며, 부분으로써 전체 대상을(예로 링거주사로 수액주사를 나타냄) 나타내거나 생식기 및 성형외과와 관련하여 완곡어구나 은유적 표현을 사용하였다. 마지막으로 소비자에 의해 생성된 신생어의 사용 및 의미 확인을 위한 전문용어의 문의

(SGOT/PT, HBs Ab 등)율이 높았다(Table 6).

(3) 매핑되지 않은 소비자 용어 분석

추출한 용어와 전문가 용어와의 매핑에서 매핑되지 않았던 주 원인은 넓은 용어의 범위, 용어의 불명확함, 건강정보 용어 경계선의 모호성, 상세한 전문용어의 사용 및 표현, 의료전문가 용어와 소비자 용어의 형태적 차이 등이었다. 전문용어의 수술 용어에서 많이 사용되는 접미사 ‘-술’의 구조는 소비자 용어에서는 접미사 ‘-수술’, ‘-술’, ‘-시술’ 구조로 다양하게

Table 6. Patterns of consumer vocabulary

Category	Example of consumer vocabulary
1. Contextual Dependency	박피 -겹질이나 가죽을 벗김 -피부 표면에 있는 흉터나 흔적들을 깎아 내어 없애는 수술
2. Use of qualitative, subjective qualifiers	심한 열, 너무 아픴, 피가 많이 남, 똥똥함
3. Misspellings	가랑이(가랭이), 궁둥이(궁뎅이, 궁뎅이), 깃스(기부스, 깃스), 꿰매다(꼬매다, 꿰매다, 꼬매다), 눈꺼풀(눈꺼플, 눈겹플), 눈두덩(눈두덩이), 뇌졸중(뇌졸중), 다래끼(눈다래끼, 눈다래끼), 두드러기(두두러기), 뒤꿈치(뒷꿈치), 메스꺼움(매스꺼움), 송곳니(송곳이), 입천장(입천정), 후유증(휴우증), 임플란트(임플랜트, 임플런트), 렌즈(렌즈)
4. Metaphor/Contraction/Description	간수치검사(간기능검사), 간수치(혈청콜레스테롤, 총단백질, 알부민, 빌리루빈, GOT/GPT효소, 알칼리인산분해효소 등 검사치), 배에서 물을 뺌(복강천자), 배란주사(배란유도제), 종아리수술(종아리퇴축술)
5. Exemplars(Use of a known members of the group or class that exemplifies the concept)	회충약(구충제), 링거주사(수액주사)
6. Colloquialisms/ Transformation	성 및 생식기- 고추(음경), 남성 기능(발기), 부부관계(성교) 성형 용어- 앞트임(내안각절제술), 뒷트임(외안각절제술), 이쁜이 수술(소음순성형술), 양귀비수술(지스팻수술) 기타- 응가(대변), 드림렌즈(각막굴절교정렌즈), 아기집(자궁), 치아뿌리(치근), 알다리(종아리 근육)
7. Neologisms	관상성형, 드림 렌즈, 사투비이(생리유도 통경제), 이쁜이수술(소음순성형술), 양귀비수술(지스팻수술)
8. Use and search of medical terminology	TSH, HDL cholesterol, HBeAg negative, chest x-ray, SGOT, UTI

Table 7. Analysis of unmapped consumer vocabulary

Cause	Example of consumer vocabulary
1. Broad Term, Lack of Specificity	치료 약물치료, 레이저 치료, 입원치료, 재활치료, 사시치료, 흰다리치료
	수술 가슴수술, 각막이식수술, 갑상선수술, 고막수술, 난시교정수술, 뇌수술, 눈수술, 위절제술, 척수수술, 코끝수술
	증상 감염 증상, 간암 말기 증상, 갑상선항진증 증상, 고혈압 증상
	검사 간초음파검사, 간염검사, 갑상선호르몬검사, 골수검사, 귀검사
	약 혈압약, 영양제, 호르몬제, 신경안정제, 관장약, 변비약, 무좀약
	물품 유도철사(스텐트), 인공관절, 보형물
	2. Intended terms not clear
3. Ambiguous domain of Health information	산성수, 성폭력
4. Specialized term, Overspecificity	GDHPQQ 방식, 포도당 PQQ방식, 밀가루 뿌린 것처럼 하얗게 됨
5. Difference of morphology between consumer vocabulary and medical vocabulary	전문 용어 담석제거술, 종양절제술, 유방축소술, 개복술, 낙태술 탈모(증), 무뇨(증), 용종절제술
	소비자 용어 담석제거수술, 종양절제수술, 유방축소수술, 개복수술, 낙태(수술) 탈모(증상), 무뇨(증상), 용종제거

사용되었고, 전문용어에서 증상 용어로 많이 사용되는 접미사 ‘-증’의 구조는 소비자 용어에서는 ‘-증상’의 구조로 많이 사용되었다. 또한 전문용어에서는 장거나 조직의 일부를 잘라내는 의미로 접미사 ‘-절제(술)’를 많이 사용하지만 소비자 용어에서는 접미사 ‘-제거’ 구조의 사용률이 높았다(Table 7).

IV. 고찰

정보화 사회는 의료전문가와 건강정보 소비자 사이의 지식격차를 감소시킬 수 있는 엄청난 잠재력을 제공하고 있기는 하지만 새로운 첨단기술로의 접근권을 가진 사람들과 그에서 제외된 사람들 사이의 격차를 더욱 벌려놓을 위험성도 가지고 있다. 이러한 문제의 가장 큰 원인은 건강정보 소비자 측에서 발생하는 장애요인으로 소비자들의 건강정보 소양이 낮아수록 건강정보를 제대로 이해하지 못하여 의사소통의 장애를 겪게 되고, 자신의 건강문제를 잘 돌보지 못하기 때문에 진료의 결과가 좋지 않게 되어 건강상태의 악화 및 입원 확률이 높아지는 것으로 나타났다. 이러한 문제를 더욱 복잡하게 만드는 요인으로는 대부분의 건강교육자료들이 국민들의 평균 독해력의 수준을 넘어서는 수준이며, 인터넷상의 건강정보는 그보다 더 높은 독해력을 요구하고 있다는 것이다⁶⁾.

소비자 용어의 확인은 소비자 용어를 잘 이해하고 가로막힌 벽을 제거하기 위한 시스템을 구성하는데 필수적인 단계이며 나아가 의료전문가 용어를 배우는데 시간이 없거나 배우려 하지 않는 사람들에게 유용하게 활용될 수 있다. 또한 정보검색 향상뿐 아니라 의사소통의 차이를 좁히기 위한 교육적 목적으로 소비자의 개념적 윤곽을 이해하는데 중요하다고 할 수 있다⁴⁾.

국외에서는 이미 2000년 이후 건강정보 소비자 용어 연구가 활발히 진행되어 의료 전문가 용어와 건강정보 소비자 용어와의 비교 분석을 통하여 두 용어간의 차이 및 건강정보 소비자 용어의 특징을 여러 연구에서 보여준 바 있다. 그러나 국내에서는 최근 들어 정부나 보험회사를 포함한 의료 관련 기관과 의료인들이 이해하기 쉬운 건강정보 소비자 용어 제공의 필요성을 인식하기 시작하여 소비자를 위한 일반 의학 용어 도서 출간 및 쉬운 용어로 행정용어와 의료용어(보험심사평가원 2007년 3월부터 2,113개 시행)를 전

환하여 관련 문서에 적용하고 있다. 그러나 건강정보 소비자 용어의 특성과 소비자들이 원하는 정보를 빠르고 정확하게 찾도록 지원하기 위한 도구 및 응용프로그램 생성을 위한 체계적인 연구는 아직 전무한 실정이다.

1. 소비자 용어

소비자 용어의 빈도분포 결과를 보면 1회(42.3%)와 2회(16.2%) 추출된 용어(우선어)가 차지하는 비율이 58.5%로 인지도가 높은 일부 용어들을 제외한 나머지 용어들의 사용빈도는 매우 낮은 것으로 나타났다. 또한 소비자들이 많이 사용하는 증상(24.0%), 치료(17.1%), 검사(23.2%) 범주의 용어들은 질병명(61.2%) 및 신체(55.2%) 범주 용어들과 비교하여 낮은 매핑률을 보였다. 이렇게 빈도가 낮고 다양하게 표현되는 용어들을 현재의 단순한 질의어 검색 중심의 구조로는 소비자가 원하는 정보를 얻기란 어려울 것이며, 소비자가 많이 사용하는 증상, 치료, 검사 범주의 용어들과 매핑률이 높은 질병명 및 신체 범주 용어와 연결이 필요하다고 사료된다.

유전/호르몬/효소/신경/세포/미세구조(85개)를 나타낸 용어 중 상당수가 검진이나 의사와의 진료시 알게 된 용어들을 문의하는 것들로 일상 경험과 관련된 신체(556개) 범주와 비교하여 매우 낮은 용어 수 및 사용빈도를 나타냈다. 이는 Blois의⁷⁾의 일상용어와 전문용어와의 분류 연구결과를 참조하면 소비자 용어는 일상용어(신체 범주- 유전/호르몬/효소/신경/세포/미세구조 제외)의 사용빈도(10,814회)가 전문용어(유전/호르몬/효소/신경/세포/미세구조 범주)의 사용빈도(384회) 보다 매우 높은 것을 반영한 것으로 보인다.

소비자 용어는 다양한 품사 및 단어, 숙어 및 문장 등의 형태로 사용되었고 주관적인 단위 사용이 높았으며 은유 및 환유, 축약, 서술적 방법으로 대상을 표현하거나 상위 넓은 범위의 용어를 사용하였다. 또한 의료 전문용어와의 형태적 차이로 의료 전문용어 중심의 용어체계에서는 소비자가 원하는 정보를 찾는 데 어려움이 있으며 불명확한 용어로 표현하거나(예로 관상동맥을 ‘심장에서 큰 동맥’으로 표현) 질병명 대신 신체의 일부를 단순히 좋지 않다고 표현하는 경우 유사형태 지원만으로는 소비자가 원하는 정보를 얻을 수 없기 때문에 소비자가 나타내고자 하는 개념

또는 의미 중심의 접근이 이루어져야 할 것이다.

2. 의료전문가 용어

본 연구의 1, 2차 통제어휘 대상 모두가 의학용어 원어에 대한 번역어 중심으로 이루어져 있고 용어 수, 일관성, 체계성에 있어서도 매핑 대상으로 충분치 않았다. 1차 통제어휘 대상으로 선정하였던 의학용어 4집의 경우 이전 용어집과 달리 우리말 표현으로 바꾸는데 역량을 기울였는데 소비자와 의료전문가 사이의 원활한 정보전달에 어려움이 있는 문제점을 개선하기 위한 측면에서는 바람직하나 소비자들이 실제 사용하는 용어와 차이가 있어 매핑 과정에서 널리 사용되는 용어들이 누락되는 단점이 있었다. 용어의 체계성 측면에서 또한 통제어휘 대상 대부분이 문제점을 갖고 있어 많이 사용되는 용어 중 예로 갑상샘약, 고혈압약은 있으나 관절염약, 당뇨약 등은 용어집에서 검색되지 않았고, 간염보균자는 없으나 하위용어인 간염보균자상태는 용어집에 포함되어 있었다. 대한약사회에서 제공하는 의약품 검색정보 또한 단순히 약품명 검색 기능으로 한정되어 정확한 약품명을 입력해야만 원하는 정보를 얻을 수 있었다.

일반 소비자들의 의료정보에 대한 관심이 점점 높아지고 있는 시점에서 통제어휘 대상으로써 의료전문 용어의 개념들에 대한 세부 영역별 용어정리 및 전문 용어사전이나 데이터베이스의 구축은 매우 중요하다고 할 수 있다.

3. 소비자 용어와 의료전문가 용어와의 매핑

소비자 용어(우선어)와 의료전문가 용어(통제어휘) 간 1차 매핑된 우선어는 30.7%, 통제어휘가 증가된 2차 매핑까지는 41.2%, 동의어 매핑까지는 46.3%로 소비자 용어와 의료전문가 용어와의 매핑률은 낮은 편이었으며 통제어휘 증가 및 동의어 확장시 매핑률의 향상을 보여주었다. 이는 추출된 용어의 49%에서 UMLS concepts과 매핑되었던 Zeng의 2002년 연구⁸⁾ 결과와 유사하며 통제어휘의 용어 수와 동의어 확장이 매핑률 향상에 영향을 준 것으로 보여진다. 또한 추출 용어와 우선어와의 매핑률 차이를 보기 위해 실시한 통제어휘와의 매핑 결과에서는 우선어가 1차 매핑에서 4.3%, 2차 매핑까지는 7.1% 높게 나타났다. 이러한 결과는 표본(전체 추출 용어)의 크기가 작았

기 때문에 우선어로 연결할 수 있는 용어의 수가 적어 낮은 차이의 매핑률을 보여주었다고 여겨진다.

4. 건강정보 소비자 용어와 의료전문가 용어 연결방안

Eysenbach는 소비자의 건강과 관련한 진단, 검사결과, 개인별 위험인자, 약처방전 정보를 대상자의 손안에 넣어줌으로써 소비자에게 힘을 주어야 한다고 하였다⁹⁾. 그러나 이를 위한 가장 큰 문제는 의료 영역 안으로 소비자의 접근성을 높이는 것이고, 그렇게 되기 위해서는 소비자가 나타내고자 하는 개념을 전달하기 위해 사용되는 용어의 형태 및 특징에 대해 먼저 명백하게 설명할 수 있어야 한다⁴⁾. 이를 위한 첫 번째 접근 방법은 소비자에 의해 인지된 용어의 형태 및 구조에 대한 규칙을 확인하고 분류하는 것이다. 두 번째 접근방법은 응용프로그램을 이용하는 것으로서 첫 번째 단계에서 확인 및 분류된 건강정보 소비자 용어의 패턴을 이용한 자동화된 용어추출 프로그램 개발을 예로 들 수 있다. 자동화된 용어추출 응용프로그램은 소비자들이 그들의 의료진에게 이메일로 문의를 하거나 상담을 할 때 본인의 증상을 보다 정확한 용어로 전할 수 있도록 도울 수 있고, 소비자가 직접 전자건강기록에 데이터를 정확하게 입력할 수 있도록 지원할 것이다. 또한, 전산시스템은 질의어나 자가증상과 같이 소비자에 의해 직접 입력된 건강 데이터를 정확하게 처리할 수 있어야 하며, 반대로 일반적이거나 개인적인 건강정보는 소비자가 이해할 수 있도록 명확하게 표현되어야 할 것이다³⁾. 특히 쌍방향 의사소통이 효과적으로 이루어지기 위해서는 소비자의 이해를 돕는 것이 중요하기 때문에 소비자에게 친근한 용어(Consumer friendly display, CFD)의 제공과 같은 건강정보 소비자 용어의 지원이 필요하다¹⁰⁾.

전문용어는 소비자가 나타내는 용어보다 검색률이 높기 때문에 소비자로부터 입력된 질의어의 자동적인 전문용어와의 매핑을 통해 건강정보 소비자 용어로서 전자건강기록 및 검사결과를 접하게 되므로 이해 및 의사소통이 증진될 수 있다. 그러나 이를 위해서 단순한 소비자 용어 목록과 사전으로는 의료통제어휘를 대체할 수는 없을 것이다. Tse의 연구에 따르면, 미국 통합의료용어 시스템인 UMLS는 건강정보 소비자 용어들을 통합하기에 적절한 곳이 될 수 있지만 증거가 반적이고 인지언어학적인 연구가 보강된 체계적인 소

비자 용어 연구가 이루어져야 한다고 하였다⁴⁾. 이를 위해서 ‘소비자 개념’, ‘의료 비전문가의 의식모델 및 의료해석모델’에 관한 연구결과는 건강정보 소비자 용어 안에 통합되어야 하고, 건강정보 소비자 용어의 특성 분석과 응용프로그램의 개발 및 적용이 이루어져야 하며, 건강정보 용어학 및 다양한 사회문화적 영향의 연구 또한 동시에 행해져야 할 것이다. 또한 실제 건강정보 소비자에 의해 사용되는 용어로부터 파생된 표현들로 구성된 전산화된 용어집은 전문가의 개념에 연결된 후 전문가와 소비자에 의해 확인 및 평가되어야 하며 소비자가 필요로 하는 수준으로 제공되어야 할 것이다. 건강정보 소비자 용어 연구를 위해서는 의사, 간호사, 정보학자뿐만 아니라 언어학자, 의학전문사서, 건강 교육자, 건강정보 소비자 등의 참여가 요구된다. 중요한 점은 건강정보 소비자 용어는 보급 및 정화, 발전을 위해 개방적 접근이 촉진되어야 하고, 소비자에 의해 사용되는 언어의 복잡성 때문에 방해 받아서는 안되며, 개별적이고 상황에 적합하게 소비자 요구에 맞추어져 소비자의 접근이 용이하도록 건강정보가 제공되어야 한다는 것이다³⁾.

본 연구에서는 표본(추출 용어의 수)의 크기가 작았기 때문에 우선어와 연결할 수 있는 용어의 수 또한 적어 추출 용어와 큰 차이를 보여주지 못했다고 여겨진다. 이에 다양한 소비자층을 대표할 수 있는 표본 크기의 보완이 필요하다. 그러나 표본의 크기가 커진다 하더라도 유사형태의 우선어와의 연결은 매핑률을 크게 높이는 데는 한계가 있어 보인다. 그보다 동의어 및 유사어 목록 형태로 구성된 우선어 목록이나 동의어 순환 목록이 제공된다면 매핑률은 보다 높아질 것이다. 그러나 이 또한 통제어휘가 간단할 때는 효율적일 수 있으나 용어의 수가 많고 복잡한 의료용어에는 적합하지 않기 때문에 동일한 관계, 계층적 관계, 연관 관계를 구분하고 시간이 지날수록 사용자의 인터페이스를 구축하고 고급화하는데 힘과 유연성을 제공하는 시소러스의 활용이 필요하다. 미국의 UMLS와 같은 한국형 통합의료용어시스템의 구축 시 통합 의료 용어의 개념들과 소비자 용어와의 매핑을 통해

소비자 용어의 형태 분석, 개념 기반 분석, 형태-개념 분석 등 개념 중심의 접근 및 분석이 필요하며, 무엇보다 소비자 용어의 특성에 대한 체계화된 연구가 선행되어야 한다. 또한 건강정보 소비자 용어를 대표하는 우선어와 의료전문가 용어와의 상이함이 여전히 존재하므로 양쪽을 연계해주는 구체적인 방안이 필요하다.

참고문헌

1. Tyson TR, Booth-Clibborn N, Oliver S. Consumer health information needs to be rigorous, complete, and relevant. *BMJ* 2000;321(7255):240.
2. Chi-Lum B. Consumers using the internet for medical information. *J Med Pract Manage* 1999;14:196-198.
3. Qing Zeng, Tony Tse. Exploring and Developing Consumer Health Vocabularies. *J Am Med Inform Assoc* 2006;13(1):24-29.
4. Anthony Y Tse. Identifying and characterizing a “consumer medical vocabulary”. University of Maryland 2003.
5. Seung-Bin Han, Jinwook Choi. Query Expansion Condensed UMLS Metathesaurus for Medical Information Retrieval. *Journal of Korean Society of Medical Informatics* 2004;10(1):43-53.
6. Jeongeun Kim. Health Informatics. Hyunmoon books 2005.
7. Blosis MS. Information and Medicine. Berkeley, CA: University of California Press 1984.
8. Zeng Q, Kogan S, Ash N, Greenes RA. Characteristics of consumer terminology for health information retrieval. *Methods of Information in Medicine* 2002;41(4): 426-434.
9. Eysenbach G. Consumer health informatics. *BMJ* 2000; 320:1713-6.
10. Zeng QT, Tse T, Crowell J, Divita G, Roth L, Browne AC. Identifying consumer-friendly display (CFD) names for health concepts. *AMIA Annu Symp Proc* 2005;859-63.