

(12) 특허협력조약에 의하여 공개된 국제출원

(19) 세계지식재산권기구
국제사무국



(43) 국제공개일
2011년 10월 6일 (06.10.2011)

PCT

(10) 국제공개번호
WO 2011/122897 A2

- (51) 국제특허분류:
G06F 17/00 (2006.01) G06F 19/00 (2011.01)
G06F 15/16 (2006.01)
- (21) 국제출원번호: PCT/KR2011/002264
- (22) 국제출원일: 2011년 4월 1일 (01.04.2011)
- (25) 출원언어: 한국어
- (26) 공개언어: 한국어
- (30) 우선권정보:
10-2010-0030059 2010년 4월 1일 (01.04.2010) KR
- (71) 출원인 (US 을(를) 제외한 모든 지정국에 대하여): 서울대학교산학협력단 (SNU R&DB FOUNDATION) [KR/KR]; 서울 관악구 신림동 산 56-1, 151-015 Seoul (KR).
- (72) 발명자: 김
- (75) 발명자/출원인 (US 에 한하여): 김홍기 (KIM, Hong Gee) [KR/KR]; 서울 관악구 봉천동 산 4-2 서울대학교수 아파트 122E-202, 151-050 Seoul (KR). 김웅희 (KIM, Eung Hee) [KR/KR]; 서울 은평구 대조동 195-20 신세게드립빌아파트 B 동 502 호, 122-842 Seoul (KR). 송승재 (SONG, Seung Jae) [KR/KR]; 서울 강남구 청담동 33 청담삼성 1 차아파트 101 동 1909 호, 135-950 Seoul (KR).

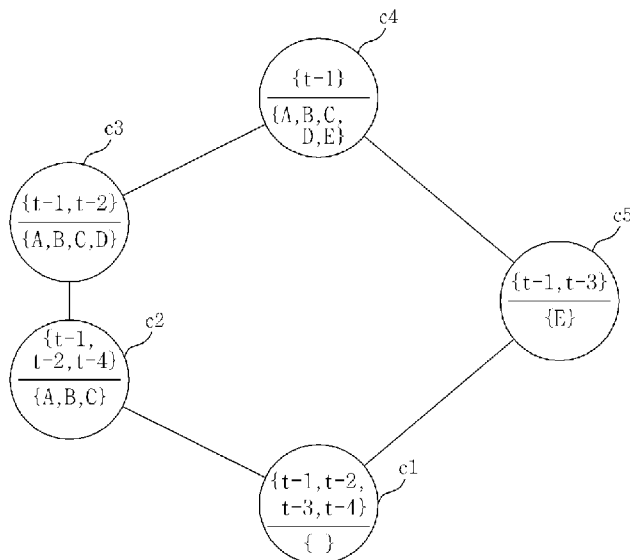
- (74) 대리인: 특허법인 정직과 특허 (HONESTY AND PATENT IP LAW FIRM); 서울특별시 강남구 역삼동 719-5 도원빌딩 8 층, 135-920 Seoul (KR).
- (81) 지정국 (별도의 표시가 없는 한, 가능한 모든 종류의 국내 권리의 보호를 위하여): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.
- (84) 지정국 (별도의 표시가 없는 한, 가능한 모든 종류의 역내 권리의 보호를 위하여): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), 유라시아 (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), 유럽 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

[다음 쪽 계속]

(54) Title: SYSTEM AND METHOD FOR SUPPORTING CONCEPT LATTICE-BASED QUERY TERM MAPPING BY WEIGHT VALUES OF USERS

(54) 발명의 명칭: 사용주체의 가중치를 이용한 개념격자 기반 질의용어 매핑지원 시스템 및 방법

[Fig. 5]



(57) Abstract: The present invention relates to a concept lattice-based query term mapping system which collects terms (terms in use) used correlatively with one another by a plurality of users, and represents terms associated with specific terms (query terms) in a graph. The system takes, as an input, terms being used by each user in sets (sets of terms), makes a concept lattice graph in which the sets of terms and users are used as nodes, and makes a graph for terms related to query terms using the concept lattice graph. The system and method of the present invention involve automatically analyzing the relationship among terms which are used correlatively with one another by users, and representing the relationship in a graph, to thereby intuitively show the relationship among terms in use.

(57) 요약서: 다수의 사용주체에 의해 서로 관련 있게 사용되는 용어(사용용어)들을 수집하여, 특정 용어(질의용어)와 관련된 용어들을 관계 그래프로 표시하는 개념격자 기반 질의용어 매핑지원 시스템에 관한 것으로서, 각 사용주체에 의해 사용되는 사용용어를 집합(용어집합)으로 입력받아 용어집합과 사용주체 집합을 노드로 구성하는 개념격자 그래프를 형성하고, 개념격자 그래프를 사용하여 질의용어에 대한 관련용어 그래프를 생성하는 구성을 마련한다. 상기와 같은 시스템 및 방법에 의하여, 사용주체에 의해 서로 관련 있게 사용되는 용어 간의 관계를 자동으로 분석하여 그래프로 표시함으로써 사용용어 사이에 내재되어 있는 관계를 직관적으로 표시할 수 있다.



공개:

- 국제조사보고서 없이 공개하며 보고서 접수 후 이를 별도 공개함 (규칙 48.2(g))

명세서

발명의 명칭: 사용주체의 가중치를 이용한 개념격자 기반 질의용어 매핑지원 시스템 및 방법

기술분야

- [1] 본 발명은 다수의 의료기관 등 사용주체에서 서로 관련 있게 사용되는 사용용어들에 대한 개념격자를 형성하여, 질의용어를 서로 관련 있게 사용되는 사용용어로 매핑하는 것을 지원하는 개념격자 기반 질의용어 매핑지원 시스템 및 방법에 관한 것이다.
- [2] 또한, 본 발명은 질의용어와 관련된 용어들을 용어노드로 표시되는 계층구조의 그래프로 구성하고, 질의용어를 루트노드로 하여 각 용어노드에 상위 용어노드의 관련성을 지지도 또는 확신도로 표시하는 개념격자 기반 질의용어 매핑지원 시스템 및 방법에 관한 것이다.
- [3] 또한, 본 발명은 사용용어 사이의 관련성을 지지도 또는 확신도로 표시하되, 상기 사용용어를 사용하는 사용주체의 등급을 반영하여 지지도 또는 확신도를 계산하는 개념격자 기반 질의용어 매핑지원 시스템 및 방법에 관한 것이다.

배경기술

- [4] 일반적으로 우리나라에서 사용되고 있는 상당수의 의학 용어가 외래어 또는 복잡한 한자어 상태로 사용되고 있다. 또한 의료인들 사이에서도 동일한 개념에 대해 여러 가지 용어를 사용함으로써 원활한 정보 소통에 문제가 발생하고 있다.
- [5] 이점 때문에 의학이 일반인에게는 이해하기 어렵고 쉽게 접근하기 어려운 학문으로 인식되고 있으며 환자와 의료인 사이의 원활한 정보 전달에도 많은 어려움이 있다. 이러한 문제와 함께 의학 용어가 표준화되어 있지 않아 병원 간의 진료정보 교류가 쉽지 않은 문제가 대두되고 있다.
- [6] 한편, 국제적으로도 의학 용어(또는 의료용어)에 대한 표준화가 많이 이루어지고 있으나, 의료분야가 가지는 복잡성으로 인해 표준화가 쉽지 않은 실정이다.
- [7] 동일한 개념(질환)을 진료과 별로 다른 의학용어를 사용하여 표현하기도 하며, 같은 의학용어를 사용하더라도 사용처에 따라 다른 의미를 가지는 경우가 있다. 예를 들어, 우리가 ‘맹장염’이라 흔히 일컫는 질병은 ‘Appendicitis’라는 의학용어가 있음에도 상황에 따라 ‘충수돌기염’, ‘충수염’, ‘막창자꼬리염’, ‘맹장염’으로 일컬어지며, ‘Myocardial Infarction(심근경색)’ 역시 상황에 따라 ‘cardiac infarction’, ‘infarction of heart’, ‘MI’, ‘Heart Attack’ 등으로 일컬어진다. 반면에 ‘주소’의 경우, 원무과에서는 ‘Address’의 의미로 사용하지만, 진료실에서는 ‘Chief Complain(주호소)’의 의미로 사용한다.
- [8] 상기와 같이 의료용어가 복잡한 것은 의료분야 자체의 복잡성에 기인한 것으로, 이러한 의료용어를 구분하고 의료진 사이의 의사소통이나 서로 다른

분야의 의사소통을 위하여 기준이 되는 참조용어체계(Reference Terminology)와 인터페이스 용어체계(Interface Terminology)가 필요하다.

- [9] 더 나아가서, 각 임상 영역에서 필요로 하는 정보를 적시에 효율적으로 수집, 분석하기 위해서는 해당 임상영역에서 필요로 하는 지식을 정의하고, 도메인 특화된 규칙을 통하여 필요 지식들을 수집하는 용어 모형을 통하여 의미적 상호운용성을 확보하는 것이 필수적이다.
- [10] 의료분야에서 사용되는 용어체계(Terminology System)란 특정 분야에서 사용되는 의학 개념의 집합이며, 개념에 대한 정의, 개념을 일컫는 용어, 개념 간의 관계 등을 포함하는 시스템이다. 의학용어 체계는 그 목적성과 특징에 따라, 참조용어(Reference terminology), 인터페이스용어(Interface terminology), 분류체계(Classification) 등으로 구분할 수 있으며, 임상현장에서 사용하는 의학용어를 로컬용어(Local Term)라 한다.
- [11] 참조용어 체계(Reference terminology)는 의학개념과 그들 간의 관계(Inter-relationships)의 정형화된 표현의 집합을 말한다. 참조용어체계에는, SNOMED CT, RxNorm, NCI Thesaurus, ICNP 등이 있다.
- [12] 인터페이스 용어 체계(Interface terminology)의 경우 의료전문가 및 임상지원 전문가가 임상문서 내 정보를 보다 수월하게 획득 및 수집하여 전달, 처리할 수 있도록 제안된 용어체계로, 의료진이 진료 기록 작성 시점에서 사용하는 모든 용어와 진료서식지 상에 있는 모든 용어들이 이에 해당한다. 인터페이스 용어체계는 특정 병원의 의료진이 특정 의학개념을 일컬을 때 사용하는 여러 형태의 용어들과 지식의 집합이라 할 수 있으며, 이는 의료진과 참조용어체계를 인터페이스 해주는 역할을 한다.
- [13] 분류체계(Classification)는 특정 목적에 맞게 미리 정해진 분화 수준에서 자료를 통합하기 위해 모든 것을 망라하는 상호 배반적인 범주, 분류의 목적은 의료서비스의 통계나 연구를 수행하는데 사용하며, ICD(International Classification of Disease), KCD(Korean Standard Classification of Disease) 등이 있다.
- [14] 의학개념에 대한 정형화된 표현들의 집합인 참조용어체계가 있다고 하더라도, 각각 병원에서 다양한 분야의 의료진이 사용하는 로컬용어는 상이할 수밖에 없기 때문에 그 차이를 매워 주기 위해 매핑이라는 작업이 필요하다.
- [15] 그러나 매핑 작업에서 가장 큰 걸림돌은 의료 현장에서 사용되는 로컬용어의 다양성이다. 즉, 한 병원 내에서도 진료과별로 의사별로 사용되는 로컬용어가 다르고 각 병원에서 관리되고 있는 로컬 용어들도 병원마다 다르다. 로컬용어는 직접 의사 등 의료진이 사용하고 있는 용어이고, 인터페이스 용어 체계는 로컬용어들을 모아서 의미를 부여하고 개념별로 정의하여 의학개념에 대한 상황정보를 탑재할 수 있도록 하는 용어체계 시스템으로서, 모두 의료현장에서 사용되는 용어이기 때문에 그 영향력이 매우 크다.
- [16] 따라서 참조용어 체계가 의료개념을 아무리 잘 표현할 수 있도록

구성되었더라도 현장에서 사용되는 의학용어와의 차이는 현저하다. 한편, 참조용어 체계를 이용하고자 하는 궁극적 목표 중 하나는 의료현장에서 난립하는 용어들 간의 의미전달을 위한 것이다.

- [17] 의료현장에서 사용되는 용어를 포괄할 수 있는 의료용어체계를 만드는 가장 좋은 방법 중 하나는 의료현장에서 사용되는 가장 지배적인 용어로 점차 통일화시켜 나가는 것이고, 이를 위해서 현재 의료현장에서 사용되는 용어들의 사용 현황을 알아야 하는 것이 매우 중요하다.
- [18] 의료현장에서 가장 많이 사용되는 용어들이 무엇인지 알게 되면, 각 의료진 또는 사용주체들은 지배적인 용어들을 점차 이용하려고 할 것이고 쓸림현상에 의해 하나 또는 소수의 최적 용어로 수렴될 수 있을 것이다.
- [19] 그러나 단순히 얼마나 많은 의료진이 사용한다는 것보다 더 중요한 것은 어떤 의미로 사용되고 있는지를 파악하는 것이다. 의료용어는 동일한 용어라도 다르게 사용될 수 있기 때문이다.
- [20] 의료현장에서 사용되는 각 용어들의 의미를 일일이 파악하는 것은 거의 불가능하므로, 의료현장 용어들의 관계를 정량적으로 분석하여 의료용어 간에 내재되어 있는 관계를 직관적으로 추출할 수 있는 자동화된 분석기술이 절실하다.

발명의 상세한 설명

기술적 과제

- [21] 본 발명의 목적은 상술한 바와 같은 문제점을 해결하기 위한 것으로, 서로 관련 있게 사용되는 의료용어 등 사용용어들로부터 용어들 간의 관계를 분석함으로써 현장에서 사용되는 사용용어 간에 내재되어 있는 관계를 직관적으로 파악할 수 있는 개념격자 기반 질의용어 매핑지원 시스템 및 방법을 제공하는 것이다.
- [22] 또, 본 발명의 목적은 다수의 사용주체에서 서로 관련 있게 사용되는 사용용어들에 대한 개념격자를 형성하여, 주어진 질의용어에 대하여 서로 관련 있게 사용되는 사용용어로 매핑하는 것을 지원하는 개념격자 기반 질의용어 매핑지원 시스템 및 방법을 제공하는 것이다.
- [23] 또, 본 발명의 목적은 질의용어와 관련된 용어들을 용어노드로 표시되는 계층구조의 그래프로 구성하고, 질의용어를 루트노드로 하여 각 용어노드에 상위 용어노드의 관련성을 지지도 또는 확신도로 표시하는 개념격자 기반 질의용어 매핑지원 시스템 및 방법을 제공하는 것이다.
- [24] 또, 본 발명의 목적은 사용용어 사이의 관련성을 지지도 또는 확신도로 표시하되, 상기 사용용어를 사용하는 사용주체의 등급을 반영하여 지지도 또는 확신도를 계산하는 개념격자 기반 질의용어 매핑지원 시스템 및 방법을 제공하는 것이다.

과제 해결 수단

- [25] 상기 목적을 달성하기 위해 본 발명은 다수의 사용주체에서 서로 관련 있게 사용되는 사용용어들에 대한 개념격자를 형성하여, 질의용어를 서로 관련 있게 사용되는 사용용어로 매핑하는 것을 지원하는 개념격자 기반 질의용어 매핑지원 시스템에 관한 것으로서, 각 사용주체에 의해 서로 관련 있게 사용되는 용어를 집합(이하 용어집합)으로 입력받는 사용용어 수집부; 용어집합과 상기 용어집합을 서로 관련 있게 사용하는 사용주체 집합을 노드(이하 개념노드)로 구성하는 개념격자 그래프를 형성하되, 하위 개념노드의 용어 집합은 상위 개념노드의 용어집합을 모두 상속하는 개념격자 형성부; 및 질의용어를 입력받아 상기 질의용어에 대하여 용어노드로 구성하는 관련용어 그래프를 생성하되, 상기 용어노드는 상기 개념격자 그래프의 개념노드 중 하나와 대응되고, 상기 용어노드와 대응되는 개념노드는 상기 질의용어를 포함하는 질의그래프 생성부를 포함하는 것을 특징으로 한다.
- [26] 또, 본 발명은 개념격자 기반 질의용어 매핑지원 시스템에 있어서, 상기 관련용어 그래프의 루트노드(이하 질의노드)는 상기 질의용어를 포함하는 최상위 개념노드와 대응되고, 자신의 용어로서 상기 질의용어를 포함하고, 상기 최상위 개념노드와 대응되지 않는 용어노드(이하 연관노드)는 자신에 대응되는 개념노드의 용어집합에서 자신의 모든 상위노드에 포함되는 용어들을 제외한 용어를 포함하고, 상기 연관노드 또는 상기 질의노드 사이의 상하위 노드 관계는 대응되는 개념노드 간의 상하위 노드 관계와 같은 것을 특징으로 한다.
- [27] 또, 본 발명은 개념격자 기반 질의용어 매핑지원 시스템에 있어서, 상기 용어노드는 자신과 대응되는 개념노드의 의료기관 집합개수를 이용하여 계산된 지지도 또는 확신도를 가지는 것을 특징으로 한다.
- [28] 또, 본 발명은 개념격자 기반 질의용어 매핑지원 시스템에 있어서, 상기 사용용어 수집부는 각 사용주체의 등급 가중치를 입력받고, 상기 지지도 또는 확신도는 상기 등급 가중치를 반영하는 것을 특징으로 한다.
- [29] 또, 본 발명은 개념격자 기반 질의용어 매핑지원 시스템에 있어서, 상기 질의노드에 대응되는 개념노드의 용어집합이 상기 질의용어 이외 다른 사용용어를 포함하면, 상기 관련용어 그래프는 상기 다른 사용용어를 포함하는 용어노드(이하 합의노드)를 상기 질의노드의 하위노드로서 구성하되, 상기 합의노드는 상기 질의노드에 대응되는 개념노드와 대응되는 것을 특징으로 한다.
- [30] 또, 본 발명은 개념격자 기반 질의용어 매핑지원 시스템에 있어서, 상기 등급가중치는 정수배이고, 각 사용주체에 대한 사용용어 집합을 상기 등급가중치 만큼 더 중복하여, 상기 지지도 또는 확신도가 구해지는 것을 특징으로 한다.
- [31] 또, 본 발명은 개념격자 기반 질의용어 매핑지원 시스템에 있어서, 상기 용어노드의 지지도 또는 확신도는 대응되는 개념노드의 사용주체들의 등급 가중치의 합의 비율로 계산되는 것을 특징으로 한다.

- [32] 또, 본 발명은 개념격자 기반 질의용어 매핑지원 시스템에 있어서, 상기 개념격자 그래프가 적어도 2개의 그래프로 분리되는 경우, 상기 지지도의 전체 사용주체는 상기 용어노드와 대응되는 개념노드를 포함하는 분리된 그래프에 포함되는 사용주체으로 한정되는 것을 특징으로 한다.
- [33] 또, 본 발명은 상기 시스템은, 질의용어를 입력받고 상기 질의용어에 대한 관련 사용용어 간의 관계를 상기 질의용어의 관련용어 그래프로 표시하고, 각 용어노드의 상위 노드와의 관계 정도를 지지도로 표시해주는 질의용어 표시부를 더 포함하는 것을 특징으로 한다.
- [34] 또, 본 발명은 개념격자 기반 질의용어 매핑지원 시스템에 있어서, 상기 질의용어 표시부는 상기 관련용어 그래프의 깊이에 따라 단계적으로 각 용어노드에 해당하는 사용용어 및 지지도 또는 확신도를 표시해주는 것을 특징으로 한다.
- [35] 또, 본 발명은 개념격자 기반 질의용어 매핑지원 시스템에 있어서, 상기 질의용어 표시부는 상기 관련용어 그래프에서 지지도 또는 확신도가 최소 지지도 또는 최소 확신도 이상인 용어노드만을 표시해주는 것을 특징으로 한다.
- [36] 또한, 본 발명은 다수의 사용주체에서 서로 관련 있게 사용되는 사용용어들에 대한 개념격자를 형성하여, 질의용어를 서로 관련 있게 사용되는 사용용어로 매핑하는 것을 지원하는 개념격자 기반 질의용어 매핑지원 방법에 관한 것으로서, (a) 각 사용주체에서 서로 관련 있게 사용되는 사용용어를 집합(이하 용어집합)으로 입력받는 단계; (b) 용어집합과 상기 용어집합을 공통적으로 사용하는 사용주체 집합을 노드(이하 개념노드)로 구성하는 개념격자 그래프를 형성하되, 하위 개념노드의 용어 집합은 상위 개념노드의 용어 집합을 모두 상속하는 단계; 및 (c) 질의용어를 입력받아 상기 질의용어에 대하여 용어노드로 구성하는 관련용어 그래프를 생성하되, 상기 용어노드는 상기 개념격자 그래프의 개념노드 중 하나와 대응되고, 상기 용어노드와 대응되는 개념노드는 상기 질의용어를 포함하는 단계를 포함하는 것을 특징으로 한다.
- [37] 또, 본 발명은 개념격자 기반 질의용어 매핑지원 방법에 있어서, 상기 용어노드는 자신과 대응되는 개념노드의 의료기관 집합개수를 이용하여 계산된 지지도 또는 확신도를 가지는 것을 특징으로 한다.
- [38] 또, 본 발명은 개념격자 기반 질의용어 매핑지원 방법에 있어서, 상기 (a)단계에서, 각 사용주체의 등급 가중치를 입력받고, 상기 지지도 또는 확신도는 상기 등급 가중치를 반영하는 것을 특징으로 한다.
- [39] 또, 본 발명은 개념격자 기반 질의용어 매핑지원 방법에 있어서, 상기 등급가중치는 정수배이고, 각 사용주체에 대한 사용용어 집합을 상기 등급가중치 만큼 더 중복하여, 상기 지지도 또는 확신도가 구해지는 것을 특징으로 한다.
- [40] 또, 본 발명은 개념격자 기반 질의용어 매핑지원 방법에 있어서, 상기 용어노드의 지지도 또는 확신도는 대응되는 개념노드의 사용주체들의 등급

가중치의 합의 비율로 계산되는 것을 특징으로 한다.

[41] 또, 본 발명은 개념격자 기반 질의용어 매핑지원 방법에 있어서, 상기 개념격자 그래프가 적어도 2개의 그래프로 분리되는 경우, 상기 지지도의 전체 사용주체는 상기 용어노드와 대응되는 개념노드를 포함하는 분리된 그래프에 포함되는 사용주체로 한정되는 것을 특징으로 한다.

[42] 또, 본 발명은 개념격자 기반 질의용어 매핑지원 방법에 있어서, 상기 방법은, (d) 질의용어를 입력받고 상기 질의용어에 대한 관련 사용용어 간의 관계를 상기 질의용어의 관련용어 그래프로 표시하고, 각 용어노드의 상위 노드와의 관계 정도를 지지도 또는 확신도로 표시해주는 질의용어 단계를 더 포함하는 것을 특징으로 한다.

[43] 또한, 본 발명은 상기 개념격자 기반 질의용어 매핑지원 방법을 수행하는 프로그램을 기록한 컴퓨터로 읽을 수 있는 기록매체에 관한 것이다.

발명의 효과

[44] 상술한 바와 같이, 본 발명에 따른 개념격자 기반 질의용어 매핑지원 시스템 및 방법에 의하면, 의료기관 등 사용주체에서 서로 관련 있게 사용되는 의료용어 등 사용용어를 취합하여 사용용어 간의 사용관계를 자동으로 분석함으로써 사용용어 간에 내재되어 있는 관계를 직관적으로 추출할 수 있는 효과가 얻어진다.

[45] 한편, 현재 의료현장에서 사용되는 용어들의 사용 현황, 즉, 사용분포를 내포하여 관련 용어들의 관계를 표시해주거나 추천해주면, 의료현장에서 각 의료진 또는 사용주체들은 지배적인 용어를 선택하여 사용할 뿐만 아니라, 자신들이 판단하기에 의료개념을 가장 잘 표현하는 용어를 선택하게 될 것이다. 또한, 이들의 선택은 현장에서 사용되는 용어들의 관계에 영향(피드백)을 줄 것이다.

[46] 따라서 시간이 지남에 따라 용어간 적자생존에 의해 하나 또는 소수의 최적 용어로 쏠림현상이 일어날 것이고, 결과적으로, 다수의 의료진들이 서로 협력 혹은 경쟁을 통하여 얻게 되는 지적 능력에 의한 결과로 얻어진 집단적 능력인 집단지성(集團知性, Collective Intelligence)에 의해 최적 용어가 결정될 것이다. 이러한 과정은 향후 사용용어 표준화에도 크게 기여할 수 있다.

[47] 또한, 본 발명에 따른 개념격자 기반 질의용어 매핑지원 시스템 및 방법에 의하면, 사용용어 사이의 관련성을 지지도 또는 확신도로 표시하되 상기 사용용어를 사용하는 사용주체의 등급을 반영하여 지지도 또는 확신도를 계산함으로써, 현장에서 각 사용주체가 미치는 영향력을 반영하여 보다 정확한 사용용어 간의 관계를 도출할 수 있는 효과가 얻어진다.

[48] 또한, 본 발명에 따른 개념격자 기반 질의용어 매핑지원 시스템 및 방법에 의하면, 각 사용주체에서 사용하고 있는 사용용어로 개념격자 그래프를 형성하고, 이를 기반으로 하여, 각 질의용어에 대한 관련용어 관계를 보다

빠르고 정확하게 도출할 수 있는 효과가 얻어진다. 즉, 사용용어의 개수와 사용주체의 수가 매우 많으므로 관련용어를 도출하는 과정이 매우 복잡하고 많은 시간이 소요될 수 있으나, 개념격자의 그래프 특성을 이용함으로써 계산량을 상당히 줄일 수 있다.

도면의 간단한 설명

- [49] 도 1은 본 발명을 실시하기 위한 전체 시스템의 구성을 도시한 도면이다.
- [50] 도 2는 본 발명의 일실시예에 따른 개념격자 기반 질의용어 매핑지원 시스템의 구성에 대한 블록도이다.
- [51] 도 3은 본 발명의 제1 실시예에 따른 사용주체에서 서로 관련 있게 사용되는 사용용어에 대한 형식문맥을 예시한 도면이다.
- [52] 도 4는 본 발명의 제1 실시예에 따라 도 3의 형식문맥으로부터 추출한 개념을 예시한 도면이다.
- [53] 도 5는 본 발명의 제1 실시예에 따른 개념격자 그래프를 예시도이다.
- [54] 도 6은 본 발명의 제1 실시예에 따른 관련용어 그래프를 예시도이다.
- [55] 도 7은 본 발명의 제1 실시예에 따라 생성된 관련용어 그래프를 이용하여 질의용어에 대한 관련용어의 지지도를 예시한 표이다.
- [56] 도 8은 본 발명의 제2 실시예에 따른 사용주체에서 서로 관련 있게 사용되는 사용용어에 대한 형식문맥을 예시한 표이다.
- [57] 도 9는 본 발명의 제2 실시예에 따른 관련용어 그래프의 예시도 및 관련용어의 지지도에 대한 표이다.
- [58] 도 10은 본 발명의 제3 실시예에 따라 적어도 2개의 다른 매핑정보를 가지는 형식문맥을 예시한 도면이다.
- [59] 도 11은 본 발명의 일실시예에 따른 개념격자 기반 질의용어 매핑지원 방법을 설명하는 흐름도이다.
- [60] 도 12는 본 발명의 일실시예에 따라 관련용어 그래프를 생성하는 알고리즘을 예시한 도면이다.

발명의 실시를 위한 형태

- [61] 이하, 본 발명의 실시를 위한 구체적인 내용을 도면에 따라서 설명한다.
- [62] 또한, 본 발명을 설명하는데 있어서 동일 부분은 동일 부호를 붙이고, 그 반복 설명은 생략한다.
- [63] 먼저, 본 발명을 실시하기 위한 전체 시스템의 구성의 예들에 대하여 도 1을 참조하여 설명한다.
- [64] 도 1a 또는 도 1b에서 보는 바와 같이, 본 발명에 따른 질의용어 매핑지원 시스템은 네트워크 상의 서버 시스템 또는 컴퓨터 단말 상의 프로그램 시스템으로 실시될 수 있다.
- [65] 도 1a와 같이, 본 발명의 실시를 위한 전체 시스템의 일례는 사용주체 단말(10)과 질의용어 매핑지원 시스템(30)으로 구성되고 서로 네트워크(20)로

연결된다. 또, 필요한 데이터를 저장하기 위한 데이터베이스(40)를 더 구비할 수 있다.

- [66] 사용주체 단말(10)은 의료인 또는 의료기관 등 사용주체가 이용하는 PC, 노트북, 넷북, PDA, 모바일 등의 통상의 컴퓨팅 단말기이다. 의료인 등은 사용주체 단말(10)을 이용하여 자신의 의료기관에서 서로 관련 있게 사용되는 다수의 사용용어들을 매핑지원 시스템(30)으로 전송하거나, 특정 용어(또는 질의용어)에 대한 관련용어들을 참조해줄 것을 매핑지원 시스템(30)에 요청한다.
- [67] 질의용어 매핑지원 시스템(30)은 통상의 서버로서 네트워크(20)에 연결되어 질의된 용어에 대하여 관련 용어들을 검색하여 관련성을 보여주는 서비스를 제공하는 서버이다. 또, 매핑지원 시스템(30)은 각 사용주체들이 서로 관련 있게 사용되는 다수의 사용용어들을 등록할 수 있는 등록서비스를 제공한다. 매핑지원 시스템(30)은 상기 각 서비스들을 인터넷 상의 웹페이지로 제공하는 웹서버 또는 웹어플리케이션 서버 등으로 구현될 수 있다.
- [68] 매핑지원 시스템(30)은 질의용어에 대하여 이와 관련된 사용용어들을 제시하면, 사용자(또는 사용주체)들은 이 제시된 결과를 참조하여 질의용어에 대한 사용용어를 매핑한다. 그러나 매핑지원 시스템(30)에서 결과로 제공되는 용어의 관련성은 용어매핑에 한정되어 이용되는 것은 아니다. 예를 들어, 사용용어를 분석하거나 하는 등의 작업에도 이용될 수 있다. 따라서 매핑을 지원하는 것에 한정되지는 않지만, 이하에서 설명의 편의를 위해 매핑을 지원하는 것으로 기재한다.
- [69] 데이터베이스(40)는 매핑지원 시스템(30)에서 필요한 데이터를 저장하는 통상의 저장매체로서, 각 사용주체로부터 수신한 사용용어로부터 형성한 개념격자 그래프, 또는 각 사용용어별로 이와 관련된 다른 관련 사용용어들을 계층구조의 관련용어 그래프 등을 저장한다.
- [70] 도 1b와 같이, 본 발명의 실시를 위한 전체 시스템의 다른 예는 컴퓨터 단말(13)에 설치되는 프로그램 형태의 질의용어 매핑지원 시스템(30)으로 구성된다. 즉, 매핑지원 시스템(30)의 각 기능들은 컴퓨터 프로그램으로 구현되어 컴퓨터 단말(13)에 설치되어, 사용자(14)에 의해 각 사용주체들이 서로 관련 있게 사용되는 다수의 사용용어들을 컴퓨터 단말(13)의 입력장치를 통해 입력받아 등록하거나, 질의용어에 대한 관련용어들을 참조해줄 것을 요청받으면 그 결과를 컴퓨터 단말(13)의 출력장치를 통해 출력한다. 한편, 매핑지원 시스템(30)에서 필요한 데이터들은 컴퓨터 단말(13)의 하드디스크 등 저장공간에 저장되어 이용된다.
- [71] 다음으로, 본 발명의 제1 실시예에 따른 개념격자 기반 질의용어 매핑지원 시스템의 구성을 도 2를 참조하여 설명한다.
- [72] 도 2에서 보는 바와 같이, 본 발명의 제1 실시예에 따른 개념격자 기반 질의용어 매핑지원 시스템(30)은 사용용어 수집부(31), 개념격자 형성부(32), 및

질의그래프 생성부(33)를 포함하여 구성된다. 추가적으로, 질의용어 표시부(32)를 더 포함하여 구성될 수도 있다. 또한, 데이터를 저장하기 위한 데이터베이스(또는 저장공간)(40)를 더 포함하여 구성될 수 있다.

[73] 먼저, 사용용어 수집부(31)는 각 사용주체으로부터 서로 관련 있게 사용되는 사용용어를 집합으로 입력받는다. 동일한 의미로 사용되는 사용용어들도 하나의 서로 관련 있게 사용하는 사용용어들이다.

[74] 도 3a에서 보는 바와 같이, 병원A, B, C는 맹장염의 의미로, 맹장염(t-1), 어펜디시티스(Appendicitis)(t-2), 및 충수돌기염(t-4)으로 3가지를 사용하고 있고, 병원D는 맹장염(t-1) 및 어펜디시티스(t-2)의 2가지만 사용하고, 병원E는 맹장염(t-1) 및 막창꼬리염(t-3)의 2가지만 사용하고 있다. 즉, 사용용어 수집부(31)는 각 사용주체으로부터 도3a와 같은 사용용어 집합을 입력받는다.

[75] 다음으로, 개념격자 형성부(32)는 용어집합과 상기 용어집합을 공통적으로 사용하는 사용주체 집합을 노드(이하 개념노드)로 표시하고, 하위 개념노드의 용어 집합은 상위 개념노드의 용어 집합을 모두 상속하는 개념격자 그래프를 형성한다.

[76] 이하에서, 개념격자 그래프를 형성하는 방법과 그 특성들을 보다 구체적으로 설명한다.

[77] 각 사용주체에 의해 서로 관련 있게 사용되는 사용용어들은 다음 정의 1과 같이 형식문맥(Formal context)으로 정의할 수 있다.

[78] [정의 1]

[79] 형식문맥 $K=(G, M, I)$ 는 사용주체 집합 G 와 용어집합 M , 그리고 G 와 M 사이의 이항관계 $I \subseteq G \times M$ 로 구성된다. 즉, G 와 M 의 원소들은 각각 해당 형식문맥의 사용주체들과 각 사용주체들이 사용할 수 있는 사용용어들을 나타낸다. 또한, 어떤 사용주체 g 가 사용용어 m 을 사용하는 경우, gIm 또는 $(g, m) \in I$ 로 나타내며, g 는 m 을 사용한다는 것을 의미한다.

[80] 도 3b에서 보는 바와 같이, 형식문맥(Formal context)은 데이터 테이블 형태로 나타낼 수 있으며, 해당 표의 행과 열의 헤드부분은 각각 형식문맥을 구성하는 사용주체들과 사용용어들로 구성된다. 또한, 데이터 테이블의 각 셀에 대해서는, 해당 셀에 관련된 사용주체와 사용용어가 이항관계 I 를 만족할 경우에는 X 표시하고, 이외의 경우에는 빈 공간으로 남겨둔다.

[81] 이와 같은 형식문맥으로부터 서로 관련 있게 사용되는 사용용어를 갖는 사용주체들을 클러스터링(Clustering)하여, 정보의 기본단위로서 연관개념들(Concepts)을 추출할 수 있다. 각 개념들은 (O, T) 와 같은 형태의 쌍(pair)으로 정의되며, 보다 정형적인 정의는 정의 2와 같다.

[82] [정의 2]

[83] 임의의 형식문맥 $K=(G, M, I)$ 에 대하여, $O \subseteq G, T \subseteq M$ 일 때, $\text{intent}(O)=T \wedge \text{extent}(T)=O$ 를 만족하는 (O, T) 를 개념(formal concept)이라고 한다.

[84] 단, $\text{intent}(O):=\{t \in M \mid \forall o \in O: (o, t) \in I\}=O', \text{extent}(T):=\{o \in G \mid \forall t \in T: (o, t) \in I\}=T'$.

- [85] 임의의 $O \subseteq G$ 에 대하여, $\text{intent}(O)$ 에 의해 O 의 모든 사용주체들이 공통적으로 갖는 용어집합을 구할 수 있다. 예를 들면, 위의 도 3b의 형식문맥에 있어서, $O = \{D, E\}$ 에 대하여, $\text{intent}(O) = \{t-1, t-2, t3\}$ 이다. 한편, 임의의 $T \subseteq M$ 에 대하여, $\text{extent}(T)$ 에 의해 T 의 사용용어들을 사용하는 사용주체들의 집합을 구할 수 있다. 예를 들면, $T = \{t-1, t-2, t-4\}$ 에 대하여, $\text{extent}(T) = \{A, B, C\}$ 이다.
- [86] 즉, 각 개념들은 (O, T) 와 같은 형태의 쌍(pair)으로 정의되며 특히, 사용주체 집합 O 는 용어집합 T 의 extent 이며, 동시에, 용어집합 T 는 사용주체 집합 O 의 intent 가 된다. 이와 같은 방법으로 형식문맥 $K = (G, M, I)$ 로부터 추출한 모든 개념들의 집합을 $B(K) := \{(O, T) \in 2^G \times 2^M \mid \text{intent}(O) = T \wedge \text{extent}(T) = O\}$ 로 나타낸다. 도 3b의 형식문맥로부터 추출한 모든 개념들 $B(K)$ 는 도4와 같다. 이하에서, c1 내지 c5는 도 3b에 표기된 개념을 나타낸다.
- [87] 위와 같이 주어진 형식문맥로부터 추출된 개념들은 extent 또는 intent 를 토대로 상위-하위개념관계를 정의 3과 4 같이 정의할 수 있다.
- [88] [정의 3]
- [89] 주어진 형식문맥의 임의의 개념 $(O1, T1), (O2, T2) \in B(K)$ 에 대하여, 상하위관계(Superconcept-Subconcept relation) $(O1, T1) \leq (O2, T2)$ 는 일종의 반순서관계(partial order relation)로서 다음과 같이 정의된다.
- [90] $(O1, T1) \leq (O2, T2) \Leftrightarrow O1 \subseteq O2 (\Leftrightarrow T1 \supseteq T2)$.
- [91] [정의 4]
- [92] 형식문맥 $K = (G, M, I)$ 에 존재하는 모든 개념 집합을 $B(K)$ 라 하고, 개념 $(X1, Y1), (X2, Y2) \in B(K)$ 에 대하여, $(X1, Y1) < (X2, Y2)$ 를 만족하고, $(X1, Y1) < (X3, Y3) < (X2, Y2)$ 를 만족하는 $(X3, Y3)$ 가 $B(K)$ 에 존재하지 않을 때, $(X1, Y1)$ 을 $(X2, Y2)$ 의 근접 하위 이웃(lower neighbor)이라 하고, $(X2, Y2)$ 를 $(X1, Y1)$ 의 근접 상위 이웃(upper neighbor)이라 하며 $(X1, Y1) \prec (X2, Y2)$ 라 표기한다.
- [93] 형식문맥 $K = (G, M, I)$ 로부터 만들어진 모든 개념들 간의 상위-하위개념관계 $\leq$ 는 일종의 반순서관계(partial order relation)에 해당하며, 개념들 사이의 상하위관계에 의해 만들어진 계층적 개념구조를 개념격자(Concept Lattice 또는 Galois Lattice)라고 부르고 $L := (B(K), E \leq)$ 과 같이 표현한다.
- [94] 도 3b에 기술된 형식문맥로부터 추출된 개념과 개념간의 상하위관계 정보를 담고 있는 개념격자를 그래프를 사용하여 도 5와 같이 표시할 수 있다. 그래프의 각 노드는 개념노드를 표시하고, 개념노드의 상위에 위치한 레이블은 개념노드의 사용용어(또는 intent)를, 하위에 위치한 레이블은 개념노드의 사용주체(또는 extent)를 각각 나타낸다. 즉, 개념격자 그래프는 용어집합과 상기 용어집합을 공통적으로 사용하는 사용주체 집합을 노드(이하 개념노드)로 표시한다.
- [95] 상기 정의 3과 4에서 보는 바와 같이, 하위 개념노드는 자신의 상위 개념노드의

사용용어들을 모두 포함한다. 즉, 하위 개념노드의 용어 집합은 상위 개념노드의 용어 집합을 모두 상속한다. 반대로 상위 개념노드는 자신의 하위 개념노드의 사용주체를 모두 포함한다.

[96] 다음으로, 형식문맥(Formal Context)에서의 용어집합(또는 속성집합) 간의 관계 및 관계에 대한 특성을 설명한다.

[97] 먼저, 용어집합 간의 함의관계는 정의 5와 같이 정의된다.

[98] [정의 5]

[99] 주어진 형식문맥 $K=(G, M, I)$ 의 임의의 두 사용용어 $Q, R \subseteq M$ 이, $\text{extent}(Q) \subseteq \text{extent}(R)$ 를 만족하는 경우, "Q는 R을 함의한다."라고 부르며, $Q \Rightarrow R$ 로 표기한다.

[100] 임의의 $Q, R \subseteq M$ 에 대하여, $\text{extent}(Q)$ 와 $\text{extent}(R)$ 에 의해서 사용용어 Q와 R을 사용하는 병원들을 각각 구할 수 있다. 예를 들어, 도 3b의 형식문맥에 있어서, $Q=\{t-4\}$, $R=\{t-1, t-2\}$ 이라 하면 $\text{extent}(Q)=\{A, B, C\}$, $\text{extent}(R)=\{A, B, C, D\}$ 를 각각 구할 수 있으며, 이는 $\text{extent}(Q) \subseteq \text{extent}(R)$ 를 만족하므로, $Q \Rightarrow R$, 즉 $\{t-4\} \Rightarrow \{t-1, t-2\}$ 가 성립함을 알 수 있다. 다시 말해, 도 3b에 주어진 형식문맥의 데이터로부터 "충수돌기염을 사용하는 병원은 맹장염과 어펜티시티스도 역시 사용한다."라는 추가적인 정보를 추출해 낼 수 있다.

[101] 형식문맥에서의 용어집합 간의 연관관계를 지지도 및 확신도에 의해 정의 6과 같이 정의된다.

[102] [정의 6]

[103] 주어진 형식문맥 $K=(G, M, I)$ 의 임의의 두 사용용어 $Q, R \subseteq M$ 이,

[104] $\frac{|\text{extent}(Q \cup R)|}{|G|} \geq \text{minsup}$ 와 $\frac{|\text{extent}(Q \cup R)|}{|\text{extent}(Q)|} \geq \text{minconf}$ 를 만족하는 경우

[105] Q는 R과 연관된다 라고 하며, $Q \rightarrow R_{\text{minsup}, \text{minconf}}$ 로 표기한다. 단 $\text{minsup}, \text{minconf} \in [0, 1]$.

[106] 정의6에서 언급된 $\frac{|\text{extent}(Q \cup R)|}{|G|}$ 와 $\frac{|\text{extent}(Q \cup R)|}{|\text{extent}(Q)|}$ 는, 각각 연관관계

$Q \rightarrow R$ 의 지지도(support)와 확신도(confidence)라고 부르며, 특히, minsup (Minimum Support: 최소 지지도)와 minconf (Minimum Confidence: 최소 확신도)는 집합 Q가 R과 갖는 연관관계를 바라보는 분석가의 주관적인 경계값(Threshold)이다. 먼저 minsup 는 Q와 R 사이에 존재하는 관계가 전체 사용주체들 중 적어도 얼마나 되는 사용주체 사이에서 성립할 경우에, 이 두 집합 간의 연관관계를 인정할 것인가에 대한 경계 값이다. 그리고 minconf 는 Q라는 사용용어를 사용하는 병원 중 적어도 얼마나 되는 병원이 R이라는 사용용어 역시 가질 때, Q는 R과 연관관계에 있다고 판단 할 것인지에 대한 경계 값이다.

[107] 예를 들어, 표 2의 context에 있어서, $Q=\{t-2\}$, $R=\{t-4\}$ 이라 하고, $\text{minsup}=0.5$,

$\text{minconf}=0.7$ 이라 하면, $\text{lexent}(\{t-2, t-4\})/|G|=0.6 \geq \text{minsup}$, $\text{lexent}(\{t-2, t-4\})/\text{lexent}(\{t-2\})=0.75 \geq \text{minconf}$ 이므로, $Q \rightarrow R$ 즉, 50%이상의 지지도와 70%이상의 확신도에서 $\{t-4\} \rightarrow \{t-2\}$ 는 성립한다고 할 수 있다. 이를 문맥상 '어펜티시티스를 사용하는 병원 중 70% 이상이 충수돌기염을 사용하는 경향을 보이며, 이러한 현상은 전체 병원 중 50% 이상의 병원으로부터 관찰된다' 라고 해석할 수 있다. 특히, $\text{minsup}=0$, $\text{minconf}=1$ 인 경우의 연관관계는 곧 함의관계를 나타내므로, 함의관계는 연관관계의 특수한 경우라 할 수 있다.

[108] 다음으로, 질의그래프 생성부(33)는 용어노드로 구성하는 관련용어 그래프를 생성하되, 상기 용어노드는 상기 개념격자 그래프의 개념노드 중 하나와 대응된다. 이때, 상기 용어노드와 대응되는 개념노드는 상기 질의용어를 포함한다. 질의용어는 사용자(또는 의료인)에 의해 질의되는 용어로서, 관련용어 그래프는 질의용어 마다 별도로 생성된다.

[109] 도 6은 도 5의 개념격자 그래프를 이용하여 생성한 관련용어 그래프로써, 도 6a는 질의용어 t-1에 대한 관련용어 그래프이고, 도 6b는 질의용어 t-2에 대한 관련용어 그래프이다.

[110] 관련용어 그래프는 개념격자 그래프를 활용하여 질의용어에 대한 함의관계 및 연관관계를 추출하고, 가중치가 부여된 비순환 유방향성 그래프(또는 트리구조의 그래프)로 가시화하여 표현되는 그래프이다. 즉, 주어진 질의 Q에 대해서, 적합한 함의관계 $Q \Rightarrow R$ 과 연관관계 $Q \rightarrow R$ minsup , minconf 를 만족하는 'R'에 관련된 제반 정보를 개념격자 그래프로부터 추출하여 표시하는 그래프이다.

[111] 관련용어 그래프에서, 루트노드(이하 질의노드)는 상의 질의용어를 포함하고 상기 질의용어를 포함하는 최상위 개념노드와 대응된다. 상기 질의노드는 정의 7에 의하여 구해진다.

[112] [정의 7]

[113] 형식문맥 $K=(G, M, I)$ 에 대한 질의 $Q \subseteq M$ 이 주어졌을 때, $Q \subseteq Y$ 를 만족하는 임의의 개념 $(X, Y) \in B(K)$ 에 대해서, $\{(X1, Y1) \in B(K) | (X, Y) \prec (X1, Y1) \wedge Q \subseteq Y1\} = \emptyset$ 인 경우, 개념 (X, Y) 을 질의 Q에 적합한 개념이라고 부르고, $(Q, (X, Y))$ 를 질의 Q에 대한 질의노드라고 정의한다.

[114] 이하에서 보듯이, 관련용어 그래프의 용어노드는 정의 7의 질의노드와 같이 "(Q, (X, Y))"의 형태로 표현된다. 이하에서, 용어노드가 (Q, (X, Y))인 경우, "용어노드는 개념노드 (X, Y)에 대응된다."라고 표현하기로 한다. 또, "용어노드의 용어는 Q이다." 또는 "용어노드는 (용어로서) Q를 포함한다."라고 표현하기로 한다. 또한, "용어노드 (Q, (X, Y))"를 간단하게 "용어노드 Q"로도 표시하기로 한다.

[115] 예를 들어 도 3b의 형식문맥에 대하여, 질의 $Q=\{t2\}$ 라 할 때, $(X, Y)=\{A, B, C\}$, $\{t-1, t-2, t-4\}=c2$ 라 하면, $Q \subseteq Y$ 는 만족하지만, $\{(X1, Y1) \in B(K) | c2 \prec (X1, Y1) \wedge Q \subseteq Y1\} = \{c3\} \neq \emptyset$ 이므로 c2는 질의 $\{t2\}$ 에 적합한 개념이라 할 수 없다.

반면 $(X, Y) = (\{A, B, C, D\}, \{t-1, t-2\}) = c_3$ 라 하면, $Q \subseteq Y$ 를 만족하고 $\{(X1, Y1) \in B(K) | c_4 \prec (X1, Y1) \wedge Q \subseteq Y1\} = \emptyset$ 이므로, 개념 c_3 을 질의 $\{t_2\}$ 에 적합한 개념이라 할 수 있으며 $(\{t_2\}, (\{A, B, C, D\}, \{t-1, t-2\}))$ 를 질의 $\{t_2\}$ 에 대한 질의노드라 할 수 있다.

[116] 도 6a에서 용어노드 $\{t-1\}$ 이 질의노드이고, 용어노드 $\{t-1\}$ 은 개념노드 $c_4 = (\{A, B, C, D, E\}, \{t-1\})$ 에 대응된다. 또, 6b에서 용어노드 $\{t-2\}$ 가 질의노드이고, 개념노드 $(\{A, B, C, D\}, \{t-1, t-2\})$ 에 대응된다.

[117] 한편, 관련용어 그래프에서 질의노드는 네모로 표시하기로 한다.

[118] 또한, 상기 질의노드에 대응되는 개념노드의 용어집합이 상기 질의용어 이외 다른 사용용어를 포함하면, 상기 관련용어 그래프는 상기 다른 사용용어를 포함하는 용어노드(이하 함의노드)를 상기 질의노드의 하위노드로서 구성하되, 상기 함의노드는 상기 질의노드에 대응되는 개념노드와 대응된다. 상기 함의노드를 다음 정의 8에 의해 구해진다.

[119] [정의 8]

[120] 형식문맥 $K = (G, M, I)$ 에서, 임의로 주어진 질의 $Q \subseteq M$ 에 대한 질의노드 $(Q, (X, Y))$ 에 대하여, $Q \neq Y \wedge \frac{|X|}{|G|} \geq \text{minsup} \in [0, 1]$ 인 경우, $(Y \setminus Q, (X, Y))$ 를 질의 Q 에

대한 함의노드라 정의한다.

[121] 함의관계를 표현하는 함의노드를 정의함에 있어 경계 값인 minsup을 반영하는 이유는, 앞서 개념격자의 설명에서 언급했듯이, 함의관계는 연관관계의 특수한 경우이기 때문이다. 즉 고유한 의미의 함의관계를 반영하는 함의노드는 minsup=0인 경우라 할 수 있다. 반면 함의관계를 표현하는 함의노드를 정의함에 있어, minconf를 반영하지 않은 이유는 다음과 같다. 질의 Q 에 적합한 개념을 $(X, Y) \in B(K)$ 라 할 때, $Q \neq Y$ 가 성립하여 함의노드가 존재할 경우, 정의2에 의하여, 항상 $\text{extent}(Q) \subseteq \text{extent}(Q \setminus Y)$, 다시 말해 $Q \Rightarrow Q \setminus Y$ 를 만족하므로, 즉 minconf=1이므로 고려하지 않는다.

[122] 예를 들어 도 3b의 형식문맥에 대하여, 질의 $Q = \{t_2\}$, minsup=0.4일 때, 질의 Q 에 적합한 개념은 c_3 이며, 질의노드는 $(\{t_2\}, c_3)$ 이다. 또한 질의 $\{t_2\}$ 가 c_3 의 intent인 $\{t_1, t_2\}$ 와 같지 않으며, $|\{A, B, C, D\}|/|G| = 0.8 \geq \text{minsup}$ 이므로, 질의 $\{t_2\}$ 에 대한 함의노드는 $(\{t_1\}, (\{A, B, C, D\}, \{t_1, t_2\}))$ 라고 할 수 있다.

[123] 도 6b에서 용어노드 $\{t-1\}$ 이 함의노드이고, 이 용어노드는 개념노드 $(\{A, B, C, D\}, \{t_1, t_2\})$ 에 대응된다.

[124] 또한, 질의용어를 포함하되 최상위가 아닌 개념노드와 대응되는 용어노드(이하 연관노드)는 자신에 대응되는 개념노드의 용어집합에서 자신의 모든 상위노드에 포함되는 용어들을 제외한 용어를 포함한다. 그리고 상기 연관노드 또는 상기 질의노드 사이의 상하위 노드 관계는 대응되는 개념노드 간의 상하위 노드 관계와 같다. 이때, 관련용어 그래프는 사용주체 집합이

적어도 하나의 원소를 가지는 개념노드에 대해서만 대응되는 용어노드로 구성한다. 구체적으로, 상기 연관노드는 정의 9에 의해 구해진다.

[125] [정의 9]

[126] 형식문맥 $K=(G, M, I)$ 에 대한 주어진 질의 $Q \subseteq M$ 에 적합한 개념이 $(X, Y) \in B(K)$ 일 때, 임의의 $(X1, Y1), (X2, Y2) \in B(K)$ 가, $(X1, Y1) < (X, Y) \wedge (X2, Y2) \leq (X, Y) \wedge (X1, Y1) \prec (X2, Y2)$ 를 만족하고, $\frac{|X1|}{|G|} \geq \text{minsup}, \frac{|X1|}{|X|} \geq \text{minconf}$ 이면, $(Y1 \setminus Y2, (X1, Y1))$ 를 질의 Q 에 대한 연관노드이다 라고 정의한다. 단 $\text{minsup}, \text{minconf} \in [0, 1]$.

[127] 주어진 질의 Q 에 적합한 개념이 (X, Y) 라면, $(X1, Y1) < (X, Y)$ 관계에 있는 개념 $(X1, Y1)$ 의 extent인 $X1$ 은, 정의 3에 의해, $Y \supset Q$ 를 사용용어로 가지면서 $Y1 \setminus Y$ 를 추가 사용용어를 사용하는 사용주체의 집합이라 할 수 있다. 이때, 정의 3(상하위관계)과 정의 4(근접하위 이웃 및 근접상위 이웃)를 활용하여, 주어진 질의 Q 와 연관관계에 있는 사용용어를 도출해 낼 수 있다. 즉, 관련용어 그래프의 용어노드 간의 상하위 노드 관계는 대응되는 (개념노드 그래프의) 개념노드 간의 상하위 노드 관계와 같다.

[128] 예를 들어 도 3b의 형식문맥에 대하여, 질의 $Q=\{t1\}$, 경계 값 $\text{minsup}=0.1$, $\text{minconf}=0.25$ 이라 하면, 질의 $\{t1\}$ 에 적합한 개념은 $c4$ 가 된다. 이때 $(X1, Y1)=c2$, $(X2, Y2)=c3$ 이라 하면, $c3 < c4$, $c2 \leq c4$, $c2 \prec c3$ 을 만족하며, $|X1|/|G|=0.6 \geq \text{minsup}$, $|X1|/|X|=0.6 \geq \text{minconf}$ 가 성립하므로, $Y1 \setminus Y2 = \{t-1, t-2, t-4\} \setminus \{t-1, t-2\} = \{t-4\}$ 가 된다. 즉, $Q=\{t1\}$ 이고, $\text{minsup}=0.1$, $\text{minconf}=0.25$ 인 경우, $(\{t-4\}, (\{A, B, C\}, \{t-1, t-2, t-4\}))$ 를 연관노드라 할 수 있다.

[129] 지금까지 정의한, 형식문맥 $K(G, M, I)$ 에 주어진 질의 $Q \subseteq M$, 각 경계 값 minsup , minconf 에 대한, 질의노드, 합의노드 및 연관노드는, $(Y^*, (X, Y))$ 와 같은 형태로 일반화하여 나타낼 수 있다(단, $(X, Y) \in B(K)$, $Y^* \subseteq Y$).

[130] 이 때, Y^* 를 용어노드의 용어(또는 레이블), (X, Y) 를 용어노드에 대응되는 개념(Concept) 또는 개념노드라 하며, 모든 노드들의 집합을 $N(KQ)_{\text{minsup}, \text{minconf}}$ 라 표기하며, 이들 노드로부터 구성되는 그래프를 관련용어 그래프라 한다. 이하에서 용어노드를 표시하는 방식 중 하나로 레이블(또는 용어노드의 용어)로 표시한다.

[131] 도 6a 및 도 6b는 각각 질의용어 $\{t1\}$ 과 $\{t2\}$ 에 대한 관련용어 그래프이고, 질의노드는 네모난 노드로, 합의노드는 마름모꼴 노드로, 연관노드는 원형 노드로 표시된다.

[132] 다음으로, 질의용어 표시부(34)는 질의용어를 입력받고 상기 질의용어에 대한 관련 사용용어 간의 관계를 상기 질의용어의 관련용어 그래프로 표시하고, 각 용어노드의 상위 노드와의 관계 정도를 지지도 또는 확신도로 표시한다. 바람직하게는, 각 용어노드에 포함된 사용용어 및 가중치를 표시해준다.

바람직하게는 가중치는 지지도 또는 확신도를 표시해준다.

- [133] 바람직하게는, 용어노드에 전체 사용주체의 개수 대비 대응되는 개념노드의 사용주체들의 개수(사용주체 집합의 원소 개수)의 비율을 지지도로 표시한다. 또, 질의노드에 대응되는 개념노드의 사용주체들의 개수 대비 자신과 대응되는 개념노드의 사용주체들의 개수의 비율을 확신도로 표시한다.

- [134] 용어노드 간의 연관관계를 정의 10에 의해 구해진다.

- [135] [정의 10]

- [136] 형식문맥 $K=(G, M, I)$ 에 주어진 질의 $Q \subseteq M$, minsup , minconf 에 관한 두 노드 $n1=(Y1^*,(X1,Y1))$, $n2=(Y2^*,(X2,Y2)) \in N(KQ)_{\text{minsup},\text{minconf}}$ 에 대하여 수학적식 1과 같은 조건을 만족하면 $n1$ 은 $n2$ 와 질의 Q 에 대해 연관관계에 있다고 정의하고 $r(n1,n2)$ 로 표기한다.

- [137] [수학적식 1]

- [138]
$$r(n_1, n_2) \Leftrightarrow \begin{cases} (X2, Y2) < (X1, Y1) \wedge Y1^* \cap Y2^* = \emptyset \wedge Y1^* \neq Q \\ \text{혹은 } (X1, Y1) = (X2, Y2) \wedge Y1^* = Q \\ \text{(함의노드가 } N(K_Q)_{\text{minsup},\text{minconf}} \text{ 내에 존재할 경우)} \\ (X2, Y2) < (X1, Y1) \wedge Y1^* \cap Y2^* = \emptyset \\ \text{(함의노드가 } N(K_Q)_{\text{minsup},\text{minconf}} \text{ 내에 존재하지 않는 경우)} \end{cases}$$

단, $R \subseteq N(KQ)_{\text{minsup},\text{minconf}} \times N(KQ)_{\text{minsup},\text{minconf}}$

- [139] 또한, 용어노드 간의 연관관계에 대한 관계 가중치는 정의 11과 같이 정의된다.

- [140] [정의 11]

- [141] 형식문맥 $K=(G, M, I)$ 에 주어진 질의 $Q \subseteq M$, minsup , minconf 에 관한 두 노드 $n1=(Y1^*,(X1,Y1))$, $n2=(Y2^*,(X2,Y2)) \in N(KQ)_{\text{minsup},\text{minconf}}$ 에 대하여, 관계 $r=(n1,n2) \in R$ 이 존재할 때, 관계 r 의 가중치를 $W(r) = \frac{(|X2|, |X2|)}{|G| \quad |Q|}$ 라 정의하며, 를

관계 r 의 지지도(support), $\frac{|X2|}{|G|}$ 를 관계 r 의 확신도(confidence)라 각각 정의한다.

- [142] 정의 11에 의한 관계 가중치인 지지도 또는 확신도를 용어관련 그래프에 각 용어노드에 표시한다. 관련용어 그래프는 트리형태의 비선형 그래프이므로 $n2$ 는 항상 하나의 관계 $r=(n1,n2)$ 만을 갖는다. 따라서 $r=(n1,n2)$ 의 가중치를 노드 $n2$ 에 기재한다.

- [143] 도 6a 또는 도 6b는 관련용어 그래프에 가중치로서 지지도를 표시하고 있다(확신도는 생략함). 즉, 용어노드 {t-1}에서 용어노드 {t-2}로의 연관관계의 지지도(또는 노드 t-2에서 지지도)는 $| \{A, B, C, D\} | / |G| = 80\%$ 이고, 용어노드 {t-1}에서 용어노드 {t-3}로의 연관관계의 지지도(또는 노드 t-3에서 지지도)는 $| \{E\} | / |G| = 20\%$ 이다.

- [144] 도 6a의 관련용어 그래프에서, 용어노드 {t-1} 표로 나타낸 것이 도 7과 같다.
- [145] 한편, 바람직하게는, 질의용어 표시부(34)는 관련용어 그래프의 깊이에 따라 단계적으로 각 용어노드에 해당하는 사용용어 및 지지도 또는 확신도를 표시해준다. 예를 들어, 도 7과 같이, 도 6a의 관련용어 그래프를 깊이 순서에 따라 순위를 두어 순차적으로 사용용어 및 지지도 또는 확신도를 표시해준다.
- [146] [정의 12]
- [147] 형식문맥 $K=(G, M, I)$, $Q \subseteq M$, $\text{minsup}, \text{minconf} \in [0, 1]$ 에 대한 용어관련그래프 $(N(KQ)_{\text{minsup}, \text{minconf}}, R)$ 에 대하여, 임의의 순서집합(Ordered set)을 $P=\{n_1, n_2, n_3, \dots, n_m\} \subseteq N(KQ)_{\text{minsup}, \text{minconf}}$ 라 할 때, $n_1 \in P$ 의 용어는 Q 이고, $1 \leq k \leq m-1$ 를 만족하는 k 에 대하여, 모든 $n_k, n_{k+1} \in P$ 에 있어, $r(n_k, n_{k+1}) \in R$ 을 만족할 때, 순서집합 P 를 관련용어 그래프의 해석 경로라 정의하며, 'n1의 용어 $\rightarrow$ n2의 용어 $\wedge \dots \wedge$ nm의 용어'이라 해석한다. 또한 해석 경로 P 의 가중치는 $r(n_{m-1}, n_m) \in R$ 의 가중치라 정의한다.
- [148] 도 6a에 표현된 관련용어 그래프의 해석 경로의 한 예를 살펴보면 다음과 같다. 용어노드는 각각 $n_1=({t-1}, ({A,B,C,D,E}, {t-1}))$, $n_2=({t-2}, ({A,B,C,D}, {t-1, t-2}))$, $n_3=({t-4}, ({A,B,C}, {t-1, t-2, t-4}))$ 이다. 이를 보다 문맥에 맞게 해석하면, '맹장염을 사용하는 사용주체는 어펜티시티스와 충수돌기염을 사용하는 경향이 있다'라고 해석할 수 있다.
- [149] 또한 수치정보인 가중치 정보를 부과하여 해석하면 '맹장염을 사용하는 병원 중 25% 이상(정확히 60%)의 병원이 어펜티시티스와 충수돌기염을 사용하는 경향이 있다.'라고 해석할 수 있다.
- [150] 다음으로, 본 발명의 제2 실시예에 따라 사용주체의 등급 가중치를 이용하여 지지도 또는 확신도를 구하는 방식을 도 8과 도 9를 참조하여 설명한다. 도 8은 제2 실시예에 따른 사용주체에서 서로 관련 있게 사용되는 사용용어에 대한 형식문맥을 예시한 표이고, 도 9는 관련용어 그래프의 예시도 및 관련용어의 지지도에 대한 표이다.
- [151] 사용용어 수집부(31)은 각 사용주체에서 서로 관련 있게 사용되는 사용용어를 집합과 더불어 각 사용주체의 등급 가중치를 입력받는다. 등급 가중치란 의료현장에서 사용주체의 영향력에 따라 부여된 것이다.
- [152] 3차원 병원에서 사용하는 사용용어는 1차 병원에서 사용되는 사용용어 보다는 의료현장에서 미치는 영향력은 더 크다. 또한, 동일한 1차 병원이라도 본원이 분원 보다 더 큰 영향력을 가지고 있다. 또한, 동일한 병원 내에서도 사용용어의 사용자에 따라 그 영향력이 다르다. 예를 들어, 의사, 수간호사, 일반 간호사 등에 따라 역시 다르다. 따라서 상기와 같은 영향력에 따라 각 사용자 또는 병원의 등급을 구분하여 가중치를 달리 정한다.
- [153] 일실시예로서, 등급 가중치는 병상(bed) 수, 병원의 차수(1,2,3차), 매핑 행위자의 직책, 전문성(specialty) 등 다양한 변수에 의해 정해진다.
- [154] 등급가중치는 관련용어 그래프의 용어노드의 지지도 또는 확신도에

반영하는데, 형식문맥을 작성할 때 해당 사용주체의 사용용어 집합을 중복시키는 방식과 지지도 또는 확신도를 구할 때 사용주체의 개수 대신 등급 가중치로 구하는 방식이 있을 수 있다.

[155] 전자의 방식의 예로서, 도 8에서 보는 바와 같이, 등급 가중치가 높은 병원 E의 사용용어 집합을 형식문맥(formal context)에 중복하여 삽입한다. 도 8에서는 추가적으로 5개의 사용용어 집합을 더 삽입하였다.

[156] 형식문맥에 중복 삽입되면, 개념격자 그래프의 개념노드의 사용용어 집합이 변경되고, 이에 따라 관련용어 그래프도 달리 생성된다. 예를 들면, 도 9a는 도 8의 형식문맥을 기반으로 생성된 질의 {t-1}에 대한 관련용어 그래프이다. 도 6a와 비교하면, 용어노드 {t-2}의 지지도는 80%에서 40%으로 감소되었고, 용어노드 {t-3}의 지지도는 20%에서 60%으로 증가되었다. 즉, 맹장염을 막창꼬리염으로 표시하는 병원은 병원 E 밖에 없으나, 병원 E의 등급 가중치가 매우 크므로, 막창꼬리염이 어펜티시티스(Appenticitis) 보다 더 관련성이 있는 사용용어로 표시된다.

[157] 따라서 질의자에게 보여주는 용어 관련성의 우선순위도 달리 표시된다. 도 9b를 도 7과 대비하여 보면, 막창꼬리염(t-3)이 어펜티시티스(t-2) 보다 더 우선적으로 표시되게 된다.

[158] 이 경우, 등급가중치는 정수배이다. 즉, 정수배 만큼 사용용어에 대한 사용용어 집합을 중복한다.

[159] 사용주체의 개수 대신 등급 가중치로 구하는 지지도 또는 확신도를 구하는 방식은 다음 정의 9-2에 의해 구해진다. 정의 9-2는 앞서 설명한 정의 9에서 지지도 및 신뢰도를 사용주체의 개수가 아닌 등급 가중치에 의해 정의된 것이다.

[160] 즉, 개념의 지지도는 전체 사용주체의 등급 가중치의 합 대비 자신의 사용주체들의 등급 가중치의 합의 비율로 계산된다. 개념의 확신도는 연관개념의 사용주체들의 등급 가중치의 합 대비 자신의 사용주체들의 등급 가중치의 합의 비율로 계산된다.

[161] [정의 9-2]

[162] 형식문맥 $K=(G, M, I)$ 에 대한 주어진 질의 $Q \subseteq M$ 에 적합한 개념이 $(X, Y) \in B(K)$ 일 때, 임의의 $(X1, Y1), (X2, Y2) \in B(K)$ 가, $(X1, Y1) < (X, Y) \wedge (X2, Y2) \leq (X, Y) \wedge (X1, Y1) \prec (X2, Y2)$ 를 만족하고, $\frac{\sum_{h \in X1} W_h}{\sum_{h \in G} W_h} \geq \text{minsup}, \frac{\sum_{h \in X1} W_h}{\sum_{h \in X} W_h} \geq \text{minconf}$ 이면, $(Y1Y2,$

$(X1, Y1))$ 를 질의Q에 대한 연관노드이다 라고 정의한다.

[163] 단 $\text{minsup}, \text{minconf} \in [0, 1]$, W_h 는 사용주체 h의 등급가중치이다.

[164] 또한, 용어노드 간의 등급가중치를 반영한 연관관계에 대한 관계 가중치는 정의 11-2과 같이 정의된다.

[165] [정의 11-2]

- [166] 형식문맥 $K=(G, M, I)$ 에 주어진 질의 $Q \subseteq M$, minsup , minconf 에 관한 두 노드 $n1=(Y1^*,(X1,Y1)), n2=(Y2^*,(X2,Y2)) \in N(KQ)$, $\text{minsup}, \text{minconf}$ 에 대하여, 관계 $r=(n1,n2) \in R$ 이 존재할 때, 관계 r 의 가중치를 $W(r) = \left(\frac{\sum_{h \in X2} W_h}{\sum_{h \in G} W_h}, \frac{\sum_{h \in X2} W_h}{\sum_{h \in Q'} W_h} \right)$ 라

정의하며, $\frac{\sum_{h \in X2} W_h}{\sum_{h \in G} W_h}$ 를 관계 r 의 지지도(support), $\frac{\sum_{h \in X2} W_h}{\sum_{h \in Q'} W_h}$ 를 관계 r 의

확신도(confidence)라 각각 정의한다. 단, W_h 는 사용주체 h 의 등급가중치이고, $Q' = \text{extent}(Q)$.

- [167] 즉, 용어노드의 지지도는 전체 사용주체의 등급 가중치의 합 대비 자신과 대응되는 개념노드의 사용주체들의 등급 가중치의 합의 비율로 계산된다. 용어노드의 확신도는 질의노드에 대응되는 개념노드의 사용주체들의 등급 가중치의 합 대비 자신과 대응되는 개념노드의 사용주체들의 등급 가중치의 합의 비율로 계산된다.
- [168] 앞서 예에서, 병원 A, B, C, D의 등급 가중치는 1로 정하고, 병원 E의 등급 가중치는 6으로 정하면, 중복삽입 방식과 동일한 결과가 도출된다.
- [169] 다음으로, 본 발명의 제3 실시예에 따라 적어도 2개의 다른 매핑정보를 가지는 형식문맥을 이용하여 관련용어 그래프를 구하는 방식을 도 10을 참조하여 설명한다.
- [170] 도 10에서 보는 바와 같이, 사용주체에서 사용하는 사용용어의 매핑정보가 적어도 2개의 다른 매핑정보로 입력되는 경우가 있다. 즉, {t-1, t-2, t-3, t-4}의 사용용어와 {t-31, t-32, t-33}의 사용용어는 서로 다를 뿐만 아니라, 이를 이용하는 사용주체 집합도 각각 {A, B, C, D, E}와 {X, Y, Z}로 각기 다르다.
- [171] 이와 같은 경우, 하나의 형식개념(또는 매핑정보)으로 입력된다하더라도 개념격자 그래프는 G1의 매핑정보와 G2의 매핑정보 별로 따로 형성된다. 이때 지지도를 구할 때 이용되는 전체 사용주체의 수, 즉, $|G|$ ({A, B, C, D, E, X, Y, Z}의 개수)를 이용한다면, G1의 개념과 G2의 개념이 다를 수 있으므로 분석의 정확도가 떨어진다.
- [172] 따라서 개념격자 그래프가 분리되어 생성되면, 분리된 개념격자 그래프 단위로 전체 사용주체의 수를 달리 정하는 것이 바람직하다. 즉, 분리된 개념격자 그래프 G1에 대해서는 $|G1|$, 즉, {A, B, C, D, E}의 개수로 이용하고, 분리된 개념격자 그래프 G2에 대해서는 $|G2|$, 즉, {X, Y, Z}의 개수로 이용한다.
- [173] 즉, 개념격자 그래프가 적어도 2개의 그래프로 분리되는 경우, 상기 지지도의 전체 사용주체는 상기 용어노드와 대응되는 개념노드를 포함하는 분리된 그래프에 포함되는 사용주체으로 한정한다.
- [174] 등급 가중치를 이용하는 경우에는 동일하게 적용될 수 있다.
- [175] 다음으로, 본 발명의 제4 실시예에 따른 개념격자 기반 질의용어 매핑지원

방법을 도 11과 도 12를 참조하여 설명한다.

- [176] 도 11에서 보는 바와 같이, 먼저, 각 사용주체에서 서로 관련 있게 사용되는 사용용어를 집합 및 각 사용주체의 등급 가중치를 입력받아(S10), 용어집합과 상기 용어집합을 공통적으로 사용하는 사용주체 집합을 노드(이하 개념노드)로 구성하는 개념격자 그래프를 형성한다(S20). 이때, 하위 개념노드의 용어집합은 상위 개념노드의 용어집합을 모두 상속한다.
- [177] 다음으로, 질의용어를 입력받아(S30), 질의용어에 대하여 용어노드로 구성하는 관련용어 그래프를 생성한다(S40). 이때, 상기 용어노드는 상기 개념격자 그래프의 개념노드 중 하나와 대응되도록 구성하고, 상기 용어노드와 대응되는 개념노드는 상기 질의용어를 포함하도록 관련용어 그래프를 구성한다. 관련용어 그래프를 생성하는 구체적인 방법은 도 12에서 보는 바와 같다.
- [178] 상기 관련용어 그래프는 다음과 같이 형성한다. 상기 관련용어 그래프의 루트노드(이하 질의노드)는 상기 질의용어를 포함하는 최상위 개념노드와 대응되고, 자신의 용어로서 상기 질의용어를 포함하도록 구성한다. 상기 최상위 개념노드와 대응되지 않는 용어노드(이하 연관노드)는 자신에 대응되는 개념노드의 용어집합에서 자신의 모든 상위노드에 포함되는 용어들을 제외한 용어를 포함한다. 그리고 상기 연관노드 또는 상기 질의노드 사이의 상하위 노드 관계는 대응되는 개념노드 간의 상하위 노드 관계와 같이 구성한다.
- [179] 또한, 상기 질의노드에 대응되는 개념노드의 용어집합이 상기 질의용어 이외 다른 사용용어를 포함하면, 상기 관련용어 그래프는 상기 다른 사용용어를 포함하는 용어노드(이하 함의노드)를 상기 질의노드의 하위노드로서 구성하되, 상기 함의노드는 상기 질의노드에 대응되는 개념노드와 대응되게 구성한다.
- [180] 또한, 상기 용어노드의 지지도는 전체 사용주체의 개수 대비 자신과 대응되는 개념노드의 사용주체 집합개수의 비율로 구해진다. 이때, 상기 지지도는 상기 등급 가중치를 반영한다. 또, 상기 용어노드의 확신도는 질의노드에 대응되는 개념노드의 사용주체들의 개수 대비 자신과 대응되는 개념노드의 사용주체들의 개수의 비율로 구해진다. 이때, 상기 확신도는 상기 등급 가중치를 반영한다.
- [181] 등급 가중치를 반영하는 방식의 일례로서, 등급가중치는 정수배이고, 각 사용주체에 대한 사용용어 집합을 상기 등급가중치 만큼 더 중복하여, 상기 지지도 또는 확신도를 구한다.
- [182] 또, 다른 일례로서, 용어노드의 지지도는 전체 사용주체의 등급 가중치의 합 대비 자신과 대응되는 개념노드의 사용주체들의 등급 가중치의 합의 비율로 계산된다. 용어노드의 확신도는 질의노드에 대응되는 개념노드의 사용주체들의 등급 가중치의 합 대비 자신과 대응되는 개념노드의 사용주체들의 등급 가중치의 합의 비율로 계산된다.
- [183] 한편, 개념격자 그래프가 적어도 2개의 그래프로 분리되는 경우, 상기 지지도의 전체 사용주체는 상기 용어노드와 대응되는 개념노드를 포함하는 분리된 그래프에 포함되는 사용주체으로 한정된다.

- [184] 그리고 질의용어를 입력받고 상기 질의용어에 대한 관련 사용용어 간의 관계를 상기 질의용어의 관련용어 그래프로 표시하고, 각 용어노드의 상위 노드와의 관계 정도를 지지도 또는 확신도로 표시한다(S50). 바람직하게는, 상기 관련용어 그래프의 깊이에 따라 단계적으로 각 용어노드에 해당하는 사용용어 및 지지도를 표시한다.
- [185] 상기 개념격자 기반 질의용어 매핑지원 방법 중 생략된 부분은 앞서 설명한 개념격자 기반 질의용어 매핑지원 시스템의 설명을 참조한다.
- [186] 이상, 본 발명자에 의해서 이루어진 발명을 상기 실시 예에 따라 구체적으로 설명하였지만, 본 발명은 상기 실시 예에 한정되는 것은 아니고, 그 요지를 이탈하지 않는 범위에서 여러 가지로 변경 가능한 것은 물론이다.

산업상 이용가능성

- [187] 본 발명은 다수의 사용주체에서 서로 관련 있게 사용되는 사용용어들에 대한 개념격자를 형성하여, 질의용어를 서로 관련 있게 사용되는 사용용어로 매핑하는 것을 지원하는 질의용어 매핑지원 시스템을 개발하는 데 적용이 가능하다.
- [188] 특히, 본 발명은 질의용어와 관련된 용어들을 용어노드로 표시되는 계층구조의 그래프로 구성하고, 질의용어를 루트노드로 하여 각 용어노드에 상위 용어노드의 관련성을 지지도 또는 확신도로 표시하는 질의용어 매핑지원 시스템을 개발하는 데 유용하다.

청구범위

[청구항 1]

다수의 사용주체에 의해 서로 관련 있게 사용되는 사용용어들에 대한 개념격자를 형성하여, 질의용어를 서로 관련 있게 사용되는 사용용어로 매핑하는 것을 지원하는 개념격자 기반 질의용어 매핑지원 시스템에 있어서,
 각 사용주체에 의해 서로 관련 있게 사용되는 용어를 집합(이하 용어집합)으로 입력받는 사용용어 수집부;
 용어집합과 상기 용어집합을 서로 관련 있게 사용하는 사용주체 집합을 노드(이하 개념노드)로 구성하는 개념격자 그래프를 형성하되, 하위 개념노드의 용어 집합은 상위 개념노드의 용어집합을 모두 상속하는 개념격자 형성부; 및
 질의용어를 입력받아 상기 질의용어에 대하여 용어노드로 구성하는 관련용어 그래프를 생성하되, 상기 용어노드는 상기 개념격자 그래프의 개념노드 중 하나와 대응되고, 상기 용어노드와 대응되는 개념노드는 상기 질의용어를 포함하는 질의그래프 생성부를 포함하는 것을 특징으로 하는 개념격자 기반 질의용어 매핑지원 시스템.

[청구항 2]

제1항에 있어서,
 상기 관련용어 그래프의 루트노드(이하 질의노드)는 상기 질의용어를 포함하는 최상위 개념노드와 대응되고, 자신의 용어로서 상기 질의용어를 포함하고,
 상기 최상위 개념노드와 대응되지 않는 용어노드(이하 연관노드)는 자신에 대응되는 개념노드의 용어집합에서 자신의 모든 상위노드에 포함되는 용어들을 제외한 용어를 포함하고,
 상기 연관노드 또는 상기 질의노드 사이의 상하위 노드 관계는 대응되는 개념노드 간의 상하위 노드 관계와 같은 것을 특징으로 하는 개념격자 기반 질의용어 매핑지원 시스템.

[청구항 3]

제2항에 있어서,
 상기 용어노드는 자신과 대응되는 개념노드의 의료기관 집합개수를 이용하여 계산된 지지도 또는 확신도를 가지는 것을 특징으로 하는 개념격자 기반 질의용어 매핑지원 시스템.

[청구항 4]

제3항에 있어서,
 상기 사용용어 수집부는 각 사용주체의 등급 가중치를 입력받고, 상기 지지도 또는 확신도는 상기 등급 가중치를 반영하는 것을 특징으로 하는 개념격자 기반 질의용어 매핑지원 시스템.

[청구항 5]

제2항 또는 제4항에 있어서,
 상기 질의노드에 대응되는 개념노드의 용어집합이 상기 질의용어

이외 다른 사용용어를 포함하면, 상기 관련용어 그래프는 상기 다른 사용용어를 포함하는 용어노드(이하 합의노드)를 상기 질의노드의 하위노드로서 구성하되, 상기 합의노드는 상기 질의노드에 대응되는 개념노드와 대응되는 것을 특징으로 하는 개념격자 기반 질의용어 매핑지원 시스템.

[청구항 6]

제4항에 있어서,
상기 등급가중치는 정수배이고, 각 사용주체에 대한 사용용어 집합을 상기 등급가중치 만큼 더 중복하여, 상기 지지도 또는 확신도가 구해지는 것을 특징으로 하는 개념격자 기반 질의용어 매핑지원 시스템.

[청구항 7]

제4항에 있어서,
상기 용어노드의 지지도 또는 확신도는 대응되는 개념노드의 사용주체들의 등급 가중치의 합의 비율로 계산되는 것을 특징으로 하는 개념격자 기반 질의용어 매핑지원 시스템.

[청구항 8]

제6항 또는 제7항에 있어서,
상기 개념격자 그래프가 적어도 2개의 그래프로 분리되는 경우, 상기 지지도의 전체 사용주체는 상기 용어노드와 대응되는 개념노드를 포함하는 분리된 그래프에 포함되는 사용주체으로 한정되는 것을 특징으로 하는 개념격자 기반 질의용어 매핑지원 시스템.

[청구항 9]

제6항 또는 제7항에 있어서, 상기 시스템은,
질의용어를 입력받고 상기 질의용어에 대한 관련 사용용어 간의 관계를 상기 질의용어의 관련용어 그래프로 표시하고, 각 용어노드의 상위 노드와의 관계 정도를 지지도 또는 확신도로 표시해주는 질의용어 표시부를 더 포함하는 것을 특징으로 하는 개념격자 기반 질의용어 매핑지원 시스템.

[청구항 10]

제9항에 있어서,
상기 질의용어 표시부는 상기 관련용어 그래프의 깊이에 따라 단계적으로 각 용어노드에 해당하는 사용용어 및 지지도 또는 확신도를 표시해주는 것을 특징으로 하는 개념격자 기반 질의용어 매핑지원 시스템.

[청구항 11]

제10항에 있어서,
상기 질의용어 표시부는 상기 관련용어 그래프에서 지지도 또는 확신도가 최소 지지도 또는 최소 확신도 이상인 용어노드만을 표시해주는 것을 특징으로 하는 개념격자 기반 질의용어 매핑지원 시스템.

[청구항 12]

다수의 사용주체에서 서로 관련 있게 사용되는 사용용어들에 대한 개념격자를 형성하여, 질의용어를 서로 관련 있게 사용되는

사용용어로 매핑하는 것을 지원하는 개념격자 기반 질의용어 매핑지원 방법에 있어서,

- (a) 각 사용주체에서 서로 관련 있게 사용되는 사용용어를 집합(이하 용어집합)으로 입력받는 단계;
- (b) 용어집합과 상기 용어집합을 공통적으로 사용하는 사용주체 집합을 노드(이하 개념노드)로 구성하는 개념격자 그래프를 형성하되, 하위 개념노드의 용어 집합은 상위 개념노드의 용어 집합을 모두 상속하는 단계; 및
- (c) 질의용어를 입력받아 상기 질의용어에 대하여 용어노드로 구성하는 관련용어 그래프를 생성하되, 상기 용어노드는 상기 개념격자 그래프의 개념노드 중 하나와 대응되고, 상기 용어노드와 대응되는 개념노드는 상기 질의용어를 포함하는 단계를 포함하는 것을 특징으로 하는 개념격자 기반 질의용어 매핑지원 방법.

[청구항 13]

제12항에 있어서,
상기 용어노드는 자신과 대응되는 개념노드의 의료기관 집합개수를 이용하여 계산된 지지도 또는 확신도를 가지는 것을 특징으로 하는 개념격자 기반 질의용어 매핑지원 방법.

[청구항 14]

제13항에 있어서,
상기 (a)단계에서, 각 사용주체의 등급 가중치를 입력받고, 상기 지지도 또는 확신도는 상기 등급 가중치를 반영하는 것을 특징으로 하는 개념격자 기반 질의용어 매핑지원 방법.

[청구항 15]

제14항에 있어서,
상기 등급가중치는 정수배이고, 각 사용주체에 대한 사용용어 집합을 상기 등급가중치 만큼 더 중복하여, 상기 지지도 또는 확신도가 구해지는 것을 특징으로 하는 개념격자 기반 질의용어 매핑지원 방법.

[청구항 16]

제14항에 있어서,
상기 용어노드의 지지도 또는 확신도는 대응되는 개념노드의 사용주체들의 등급 가중치의 합의 비율로 계산되는 것을 특징으로 하는 개념격자 기반 질의용어 매핑지원 방법.

[청구항 17]

제15항 또는 제16항에 있어서,
상기 개념격자 그래프가 적어도 2개의 그래프로 분리되는 경우, 상기 지지도 또는 확신도의 전체 사용주체는 상기 용어노드와 대응되는 개념노드를 포함하는 분리된 그래프에 포함되는 사용주체으로 한정되는 것을 특징으로 하는 개념격자 기반 질의용어 매핑지원 방법.

[청구항 18]

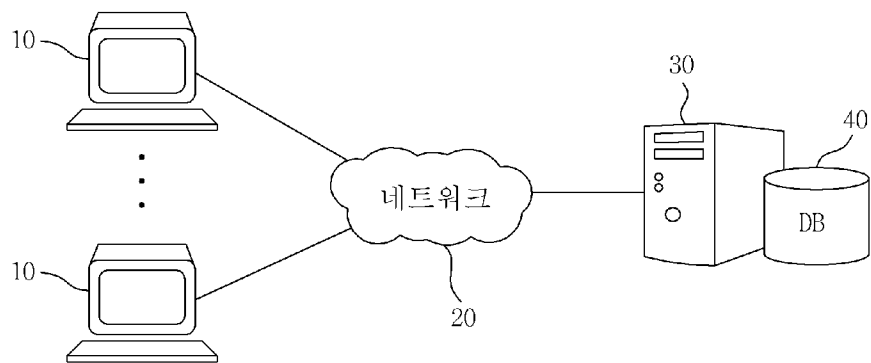
제15항 또는 제16항에 있어서, 상기 방법은,

(d) 질의용어를 입력받고 상기 질의용어에 대한 관련 사용용어 간의 관계를 상기 질의용어의 관련용어 그래프로 표시하고, 각 용어노드의 상위 노드와의 관계 정도를 지지도 또는 확신도로 표시해주는 질의용어 단계를 더 포함하는 것을 특징으로 하는 개념격자 기반 질의용어 매핑지원 방법.

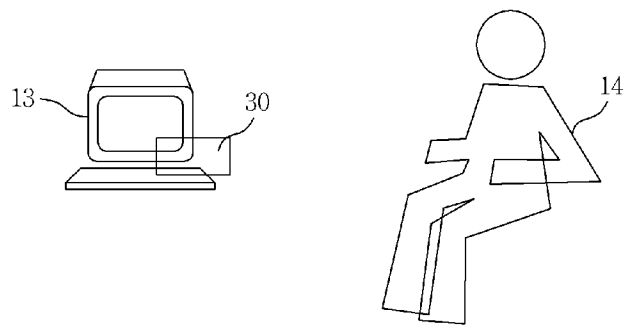
[청구항 19]

제12항 내지 제16항 중 어느 한 항의 개념격자 기반 질의용어 매핑지원 방법을 수행하는 프로그램을 기록한 컴퓨터로 읽을 수 있는 기록매체.

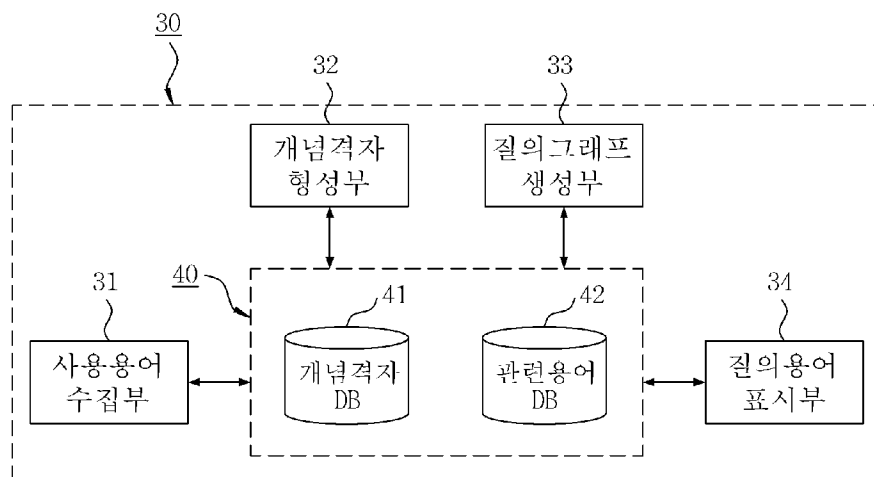
[Fig. 1a]



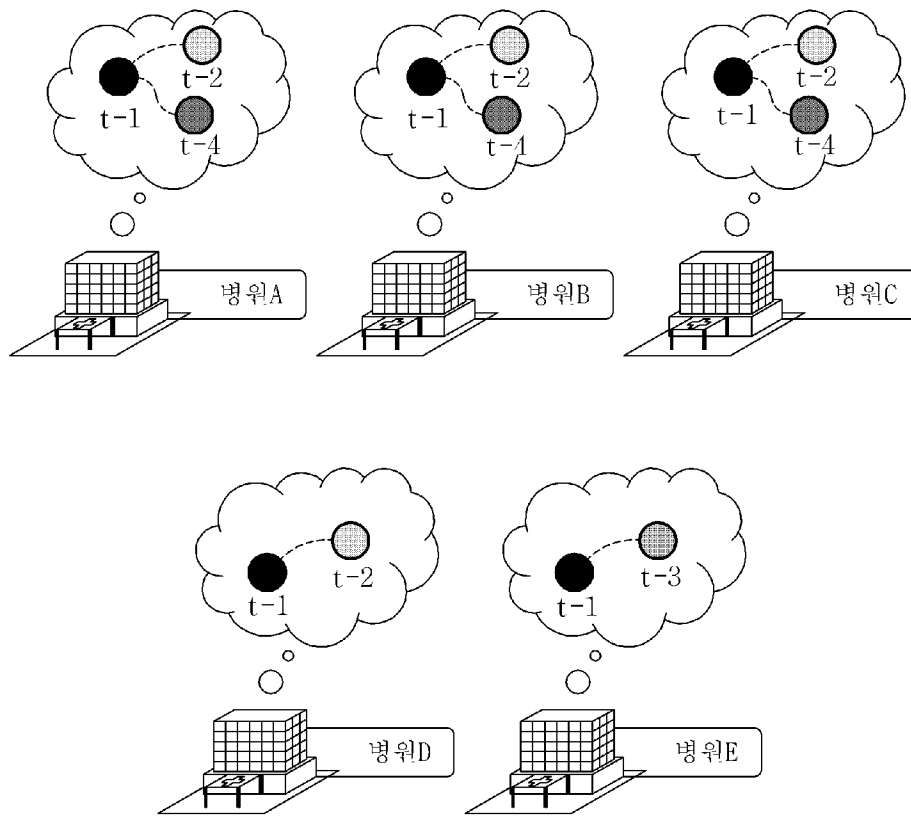
[Fig. 1b]



[Fig. 2]



[Fig. 3a]



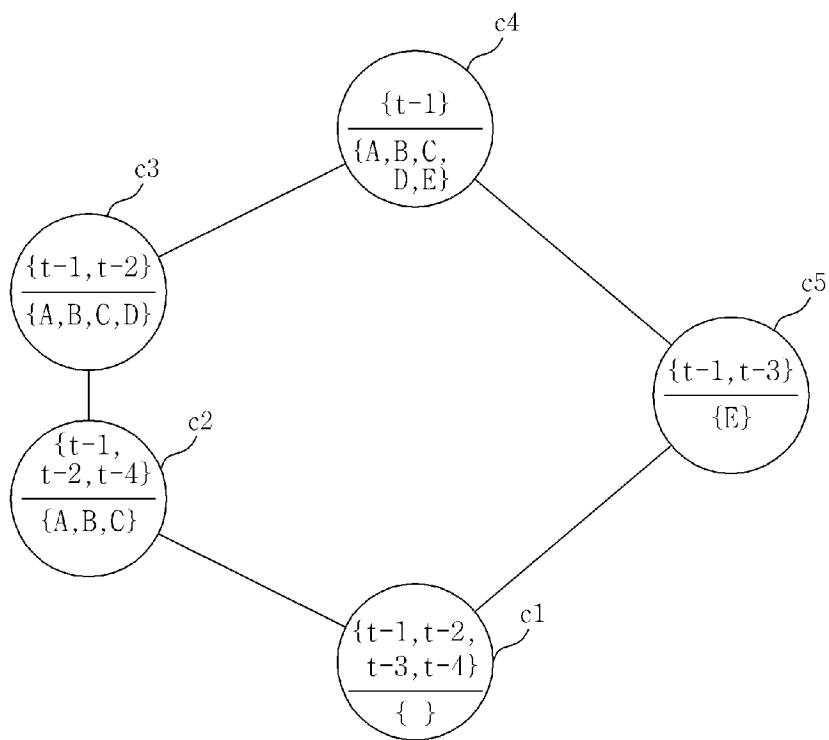
[Fig. 3b]

	t-1 (맹장염)	t-2 (Appenticitis)	t-3 (막창꼬리염)	t-4 (충수돌기염)
A	×	×		×
B	×	×		×
C	×	×		×
D	×	×		
E	×		×	

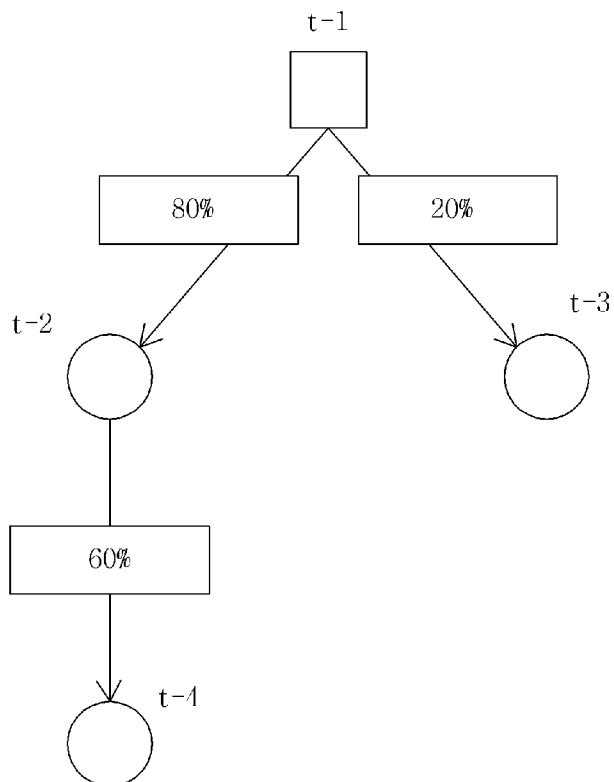
[Fig. 4]

개념	Extent	Intent
c1	$\emptyset$	$\{t-1, t-2, t-3, t-4\}$
c2	$\{A, B, C\}$	$\{t-1, t-2, t-4\}$
c3	$\{A, B, C, D\}$	$\{t-1, t-2\}$
c4	$\{A, B, C, D, E\}$	$\{t-1\}$
c5	$\{E\}$	$\{t-1, t-3\}$

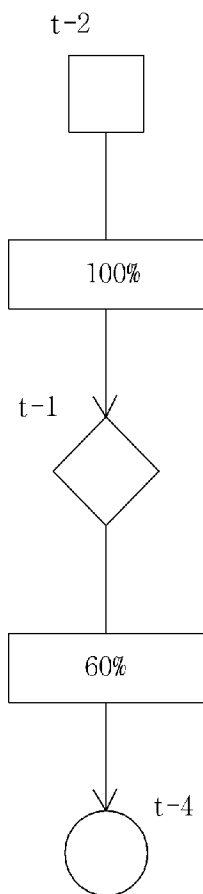
[Fig. 5]



[Fig. 6a]



[Fig. 6b]



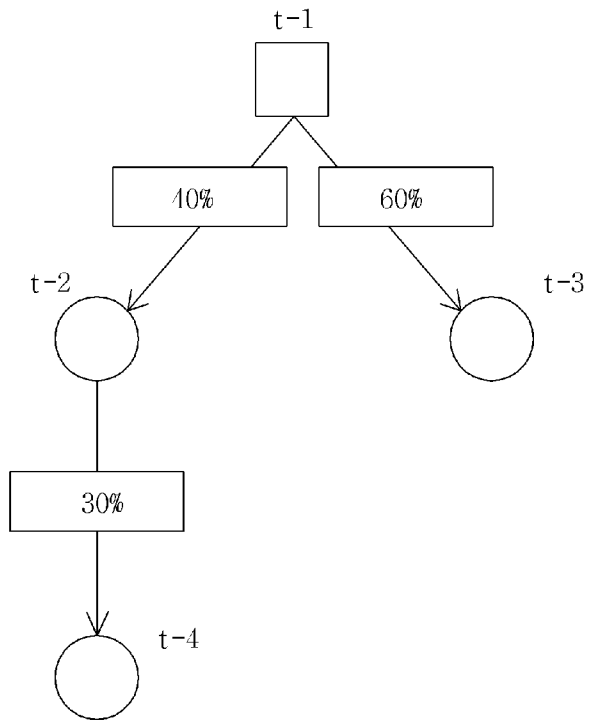
[Fig. 7]

순위	관련용어	중요도
1	t-2	80%
2	t-3	20%
3	t-4	60%

[Fig. 8]

	t-1	t-2	t-3	t-4
A	X	X		X
B	X	X		X
C	X	X		X
D	X	X		
E	X		X	
E-1	X		X	
E-2	X		X	
E-3	X		X	
E-4	X		X	
E-5	X		X	

[Fig. 9a]



[Fig. 9b]

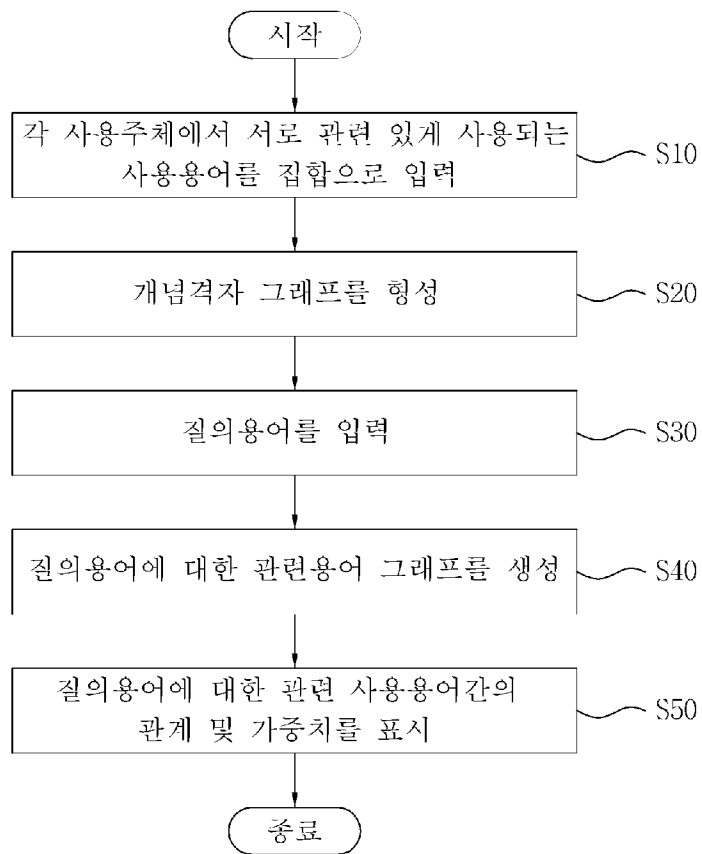
순위	관련용어	중요도
1	t-3	60%
2	t-2	40%
3	t-4	30%

[Fig. 10]

	t-1	t-2	t-3	t-4	...	t-31	t-32	t-33	...
A	X	X		X	...				
B	X	X		X	...				
C	X	X		X	...				
D	X	X			...				
E	X		X		...				
...	...	...	...	...	...	...	...	...	...

X					X		X	
Y					X	X		
Z						X	X	

[Fig. 11]



[Fig. 12]

Input: 형식분맥 $K=(G, M, I)$ 의 개념격자 $L=(B(K), E_{\leq})$, 사용자 질의 $Q \subseteq M$,
 $\text{minsup}, \text{minconf} \in [0, 1]$
Output: 용어관련그래프 $(N(KQ), \text{minsup}, \text{minconf}, R)$

1. $N(KQ), \text{minsup}, \text{minconf} := \emptyset, R := \emptyset, \text{nextLevel} := \emptyset$
2. for each $(X, Y) \in B(K)$
3. if $Q \subseteq Y$ then
4. $\text{Super} := \text{Find all } (X_1, Y_1) \in C \text{ such that } (X, Y) (X_1, Y_1) \text{ in } E_{\leq}$
5. if $Q \subseteq Y$ then
6. $N(KQ), \text{minsup}, \text{minconf} := \{(Q, (X, Y))\}$
7. if $|X|/|G| \geq \text{minsup}$ then
8. if $Y \setminus Q \neq \emptyset$ and then
9. $N(KQ), \text{minsup}, \text{minconf} := N(KQ), \text{minsup}, \text{minconf} \cup \{(Y \setminus Q, (X, Y))\}$
10. $\text{nextLevel} := \{(Y \setminus Q, (X, Y))\}$
11. Add edge $(Q, (X, Y)) \rightarrow (Y \setminus Q, (X, Y))$ to R
12. $(Q, (X, Y)) \rightarrow (Y \setminus Q, (X, Y)).\text{weight} = (|X|/|G|, 1)$
13. else then
14. $\text{nextLevel} := \{(Q, (X, Y))\}$
15. break loop;
16. if $\text{nextLevel} \neq \emptyset$ then
17. addAssociationNode(nextLevel)

function addAssociationNode(nextLevel)

1. $\text{currentLevel} := \text{nextLevel}$
2. $\text{nextLevel} := \emptyset$
3. for each $(Y^*, (X, Y)) \in \text{currentLevel}$
4. $\text{Sub} := \text{Find all } (X_1, Y_1) \in C \text{ such that } (X_1, Y_1) (X, Y) \text{ in } E_{\leq}$
5. for each $(X_1, Y_1) \in \text{Sub}$
6. if $|X_1|/|G| \geq \text{minsup}$ and $|X_1|/|Q'| \geq \text{minconf}$ then
7. if $(\{Y_1 \setminus Y\}, (X_1, Y_1)) \in N(KQ), \text{minsup}, \text{minconf}$ then
8. $N(KQ), \text{minsup}, \text{minconf} := N(KQ), \text{minsup}, \text{minconf} \cup \{(\{Y_1 \setminus Y\}, (X_1, Y_1))\}$
9. $\text{nextLevel} := \text{nextLevel} \cup \{(\{Y_1 \setminus Y\}, (X_1, Y_1))\}$
10. Add edge $(Y^*, (X, Y)) \rightarrow (\{Y_1 \setminus Y\}, (X_1, Y_1))$ to R
11. $(Y^*, (X, Y)) \rightarrow (\{Y_1 \setminus Y\}, (X_1, Y_1)).\text{weight} = (|X_1|/|G|, |X_1|/|Q'|)$
12. if $\text{nextLevel} \neq \emptyset$ then
13. addAssociationNode(nextLevel)