

(12) 특허협력조약에 의하여 공개된 국제출원

(19) 세계지식재산권기구
국제사무국



(10) 국제공개번호

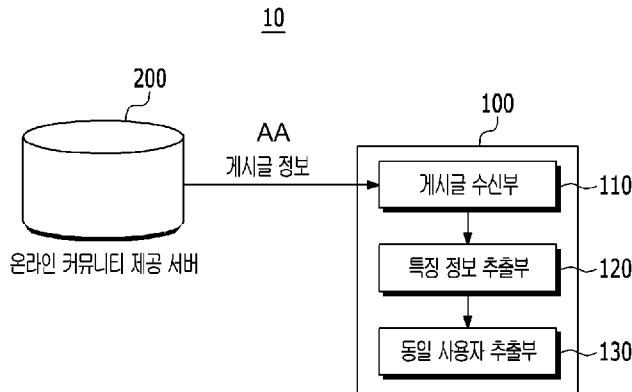
(43) 국제공개일
2018 년 6 월 14 일 (14.06.2018) **WIPO | PCT**

WO 2018/105979 A1

- (51) 국제특허분류: **G06Q 50/30** (2012.01) **G06F 17/00** (2006.01)
G06F 17/26 (2006.01) **Sanghyun**; 06941 서울시 동작구 알마타길 6, 102동 2002호, Seoul (KR).
- (21) 국제출원번호: PCT/KR2017/014108 (74) 대리인: 유미특허법인 (YOU ME PATENT AND LAW FIRM); 06134 서울시 강남구 테헤란로 115, Seoul (KR).
- (22) 국제출원일: 2017 년 12 월 4 일 (04.12.2017) (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, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, 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.
- (25) 출원언어: 한국어
- (26) 공개언어: 한국어
- (30) 우선권정보: 10-2016-0166144 2016 년 12 월 7 일 (07.12.2016) KR
- (71) 출원인: 서강대학교 산학협력단 (SOGANG UNIVERSITY RESEARCH FOUNDATION) [KR/KR]; 04107 서울시 마포구 백범로 35, Seoul (KR).
- (72) 발명자: 박석 (PARK, Seog); 03009 서울시 종로구 평창12길 16-6, 2동 202호, Seoul (KR). 박상현 (PARK, Sanghyun); 06941 서울시 동작구 알마타길 6, 102동 2002호, Seoul (KR).
- (84) 지정국 (별도의 표시가 없는 한, 가능한 모든 종류의 국내 권리의 보호를 위하여): ARIPO (BW, GH, GM, KE,

(54) Title: APPARATUS AND METHOD FOR EXTRACTING NICKNAME LIST OF SAME USER FROM ONLINE COMMUNITY

(54) 발명의 명칭: 온라인 커뮤니티에서 동일 사용자의 닉네임 목록을 추출하는 장치 및 방법



110 ... Posting receiving unit
120 ... Characteristic information extracting unit
130 ... Same user extracting unit
200 ... Online community providing server
AA ... Posting information

(57) Abstract: A method for extracting nicknames of the same user by a device operated by at least one processor comprises the steps of: receiving, from a server, a posting uploaded in an online community; extracting, from the posting, at least one piece of characteristic information that identifies a user who wrote the posting; and extracting, from a plurality of nickname sets, a nickname having a degree of similarity to the characteristic information that is equal to or higher than a predetermined reference level, as the nickname of the same user.

(57) 요약서: 본 발명은 적어도 하나의 프로세서에 의해 동작하는 장치가 동일 사용자의 닉네임을 추출하는 방법으로서, 서버로부터 온라인 커뮤니티 내에 업로드 된 게시글을 수신하는 단계, 상기 게시글에서 상기 게시글을 작성한 사용자를 식별하는 적어도 하나 이상의 특징 정보를 추출하는 단계, 그리고 복수의 닉네임 집합에서 상기 특징 정보의 유사 정도가 일정 기준 이상인 닉네임을 동일 사용자의 닉네임으로 추출하는 단계를 포함한다.

[다음 쪽 계속]

WO 2018/105979 A1



LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 유라시아 (AM, AZ, BY, KG, KZ, 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, KM, ML, MR, NE, SN, TD, TG).

공개:

— 국제조사보고서와 함께 (조약 제21조(3))

명세서

발명의 명칭: 온라인 커뮤니티에서 동일 사용자의 닉네임 목록을 추출하는 장치 및 방법

기술분야

- [1] 본 발명은 온라인 커뮤니티에서 동일 사용자의 닉네임 목록을 추출하는 장치 및 방법에 관한 것이다.

배경기술

- [2] 모바일 및 유무선 네트워크 기술의 발전에 따라 시간과 공간의 제약 없이 웹에 접근할 수 있게 되어 다양한 서비스들이 등장하고 있다. 대표적인 온라인 서비스로는 소셜 네트워크 서비스(Social Network Service, SNS), 그리고 온라인 커뮤니티(Online Community)가 있다.
- [3] 소셜 네트워크 서비스는 사용자들이 정보를 공유하고, 다른 사용자들과 의사소통할 수 있는 온라인 서비스로서, 페이스북, 트위터, 인스타그램 등이 있다.
- [4] 온라인 커뮤니티는 사용자들이 웹상에서 공통된 관심사 및 취향에 대한 정보를 생산, 공유하고 이들이 모여 활동할 수 있도록 공간을 마련해 주는 서비스로서, 웹의 초창기부터 제공된 형태의 서비스이다.
- [5] 이러한 온라인 소셜 네트워크 서비스의 확산에 따라 온라인 소셜 네트워크 사용자들이 자발적으로 업로드한 다량의 글, 이미지, 위치 태그들을 이용하여 사회 현상을 분석하고, 선거 결과를 예측하는 등 다양한 분석 자료로 활용하는 연구가 지속되고 있다.
- [6] 그러나 온라인 커뮤니티는 개인의 정보가 명확히 드러나지 않고, 닉네임을 변경함에 따라 과거의 신원과의 연속성이 단절되어 개인 정보를 식별하기 어려워 온라인 커뮤니티 내의 게시글을 비롯한 데이터들은 다양한 분석 자료로 활용되는 사례가 극히 드물다.
- [7] 온라인 커뮤니티는 익명성을 보장하기 때문에 사용자들은 더욱 솔직한 의견을 게시할 수 있으므로, 온라인 커뮤니티에 포함된 다수의 데이터를 다양한 분석 자료로 활용할 수 있도록 개인 정보를 식별하기 위한 연구가 필요하다.

발명의 상세한 설명

기술적 과제

- [8] 본 발명이 해결하고자 하는 과제는 익명성이 보장된 온라인 커뮤니티에서 사용자가 닉네임을 변경하더라도 변경 전 닉네임과 변경 후 닉네임이 동일인임을 식별할 수 있도록 동일 사용자의 닉네임 목록을 추출하는 장치 및 방법을 제공하는 것이다.

과제 해결 수단

- [9] 본 발명의 한 실시예에 따른 적어도 하나의 프로세서에 의해 동작하는 장치가

동일 사용자의 닉네임을 추출하는 방법은 서버로부터 온라인 커뮤니티 내에 업로드 된 게시글을 수신하는 단계, 상기 게시글에서 상기 게시글을 작성한 사용자를 식별하는 적어도 하나 이상의 특징 정보를 추출하는 단계, 그리고 복수의 닉네임 집합에서 상기 특징 정보의 유사 정도가 일정 기준 이상인 닉네임을 동일 사용자의 닉네임으로 추출하는 단계를 포함한다.

- [10] 상기 복수의 닉네임 집합에서 동일 사용자가 아닌 것으로 확정되는 닉네임을 제외하여 후보 닉네임 집합을 선별하는 단계를 더 포함하고, 상기 닉네임 추출 단계는 상기 후보 닉네임 집합에서 상기 특징 정보의 유사 정도가 일정 기준 이상인 닉네임을 동일 사용자의 닉네임으로 추출할 수 있다.
- [11] 상기 동일 사용자가 아닌 것으로 확정되는 닉네임은 사용 기간이 중첩되는 닉네임인 닉네임일 수 있다.
- [12] 상기 특징 정보를 추출하는 단계는 상기 사용자의 관심 정보를 식별할 수 있는 단어 특징 정보, 상기 사용자의 활동 시간 특징 정보, 상기 게시글 특징 정보, 그리고 상기 사용자와 다른 사용자와의 커뮤니케이션 관계 특징 정보 중 적어도 하나 이상의 특징 정보를 추출할 수 있다.
- [13] 상기 단어 특징 정보는 상기 게시글 텍스트에서 형태소 분석을 이용하여 적어도 하나 이상의 특징 단어, 그리고 상기 특징 단어의 사용 빈도 수 정보를 포함할 수 있다.
- [14] 상기 커뮤니케이션 관계 특징 정보는 상기 사용자가 작성한 적어도 하나의 게시글에 대하여 답글을 작성하거나 공감을 표현한 닉네임 정보인 제1 닉네임 특징 정보, 상기 사용자가 답글을 작성하거나, 공감을 표현한 타 사용자의 게시글 작성자 닉네임 정보인 제2 닉네임 특징 정보, 그리고 상기 제1 닉네임 특징 정보와 상기 제2 닉네임 특징 정보를 조합한 특징 정보 중 적어도 하나 이상일 수 있다.
- [15] 본 발명의 한 실시예에 따른 적어도 하나의 프로세서에 의해 동작하는 장치가 동일 사용자의 닉네임을 추출하는 방법은 서버로부터 온라인 커뮤니티 내에 업로드 된 게시글을 수신하는 단계, 상기 게시글에서 사용자의 관심 정보를 식별할 수 있는 단어 특징 정보, 상기 사용자의 활동 시간 특징 정보, 상기 게시글 특징 정보, 그리고 상기 사용자와 다른 사용자와의 커뮤니케이션 관계 특징 정보 중 적어도 둘 이상을 포함하는 특징 정보를 추출하는 단계, 복수의 닉네임 간 상기 특징 정보들의 유사 정도를 계산하는 단계, 상기 특징 정보마다 서로 다른 가중값을 상기 유사 정도에 부여하는 단계, 그리고 상기 유사 정도가 일정 기준 이상인 닉네임을 동일 사용자의 닉네임으로 추출하는 단계를 포함한다.
- [16] 상기 복수의 닉네임에서 동일 사용자가 아닌 것으로 확정되는 닉네임을 제외하여 후보 닉네임 집합을 선별하는 단계를 더 포함하고, 상기 닉네임 추출 단계는 상기 후보 닉네임 집합에서 상기 유사 정도가 일정 기준 이상인 닉네임을 동일 사용자의 닉네임으로 추출할 수 있다.
- [17] 상기 동일 사용자가 아닌 것으로 확정되는 닉네임은 사용 기간이 중첩되는

닉네임일 수 있다.

- [18] 상기 가중값은 유전 알고리즘을 통해 산출된 값일 수 있다.
- [19] 본 발명의 한 실시예에 따른 동일한 사용자의 닉네임을 추출하기 위한 장치는 온라인 커뮤니티에 업로드된 적어도 하나의 게시글을 수신하는 게시글 수신부, 상기 게시글에서 상기 게시글을 작성한 사용자를 식별하는 적어도 하나 이상의 특징 정보를 추출하는 특징 정보 추출부, 그리고 상기 특징 정보를 기초로 복수의 닉네임 집합에서 상기 특징 정보의 유사 정도가 일정 기준 이상인 닉네임을 동일 사용자의 닉네임으로 추출하는 동일 사용자 추출부를 포함한다.
- [20] 상기 동일 사용자 추출부는 상기 복수의 닉네임 집합에서 동일 사용자가 아닌 것으로 확정되는 닉네임을 제외하여 후보 닉네임 집합을 선별하고, 상기 후보 닉네임 집합에서 상기 특징 정보의 유사 정도가 일정 기준 이상인 닉네임을 동일 사용자의 닉네임으로 추출할 수 있다.
- [21] 상기 동일 사용자가 아닌 것으로 확정되는 닉네임은 사용 기간이 중첩되는 닉네임일 수 있다.
- [22] 상기 특징 정보 추출부는 상기 사용자의 관심 정보를 식별할 수 있는 단어 특징 정보, 상기 사용자의 활동 시간 특징 정보, 상기 게시글 특징 정보, 그리고 상기 사용자와 다른 사용자와의 커뮤니케이션 관계 특징 정보 중 적어도 하나 이상의 특징 정보를 추출할 수 있다.
- [23] 상기 단어 특징 정보는 상기 게시글 텍스트에서 형태소 분석을 이용하여 적어도 하나 이상의 특징 단어, 그리고 상기 특징 단어의 사용 빈도수 정보를 포함할 수 있다.
- [24] 상기 커뮤니케이션 관계 특징 정보는 상기 사용자가 작성한 적어도 하나의 게시글에 대하여 답글을 작성하거나 공감을 표현한 닉네임 정보인 제1 닉네임 특징 정보, 상기 사용자가 답글을 작성하거나, 공감을 표현한 타 사용자의 게시글 작성자 닉네임 정보인 제2 닉네임 특징 정보, 그리고 상기 제1 닉네임 특징 정보와 상기 제2 닉네임 특징 정보를 조합한 특징 정보 중 적어도 하나 이상일 수 있다.
- [25] 본 발명의 한 실시예에 따른 동일한 사용자의 닉네임을 추출하기 위한 장치는 온라인 커뮤니티에 업로드된 적어도 하나의 게시글을 수신하는 게시글 수신부, 상기 게시글에서 사용자의 관심 정보를 식별할 수 있는 단어 특징 정보, 상기 사용자의 활동 시간 특징 정보, 상기 게시글 특징 정보, 그리고 상기 사용자와 다른 사용자와의 커뮤니케이션 관계 특징 정보 중 적어도 둘 이상을 포함하는 특징 정보를 추출하는 특징 정보 추출부, 그리고 복수의 닉네임 간 상기 특징 정보들의 유사 정도를 계산하고, 상기 특징 정보마다 서로 다른 가중값을 상기 유사 정도에 부여하며, 상기 유사 정도가 일정 기준 이상인 닉네임을 동일 사용자의 닉네임으로 추출하는 동일 사용자 추출부를 포함한다.
- [26] 상기 동일 사용자 추출부는 상기 복수의 닉네임에서 동일 사용자가 아닌 것으로 확정되는 닉네임을 제외하여 후보 닉네임 집합을 선별하고, 상기 후보

닉네임 집합에서 상기 유사 정도가 일정 기준 이상인 닉네임을 동일 사용자의 닉네임으로 추출할 수 있다.

[27] 상기 동일 사용자가 아닌 것으로 확정되는 닉네임은 사용 기간이 중첩되는 닉네임일 수 있다.

[28] 상기 가중값은 유전 알고리즘을 통해 산출된 값일 수 있다.

발명의 효과

[29] 본 발명의 실시예에 따르면 온라인 커뮤니티에서 사용자가 닉네임을 변경하더라도 동일 사용자의 닉네임 목록을 추출함으로써 서로 다른 닉네임으로 활동한 데이터를 일인(一人)의 자료로 수집할 수 있다.

[30] 본 발명의 실시예에 따르면 익명성이 보장된 온라인 커뮤니티에서 어뷰징(abusing)을 일삼는 사용자를 차단할 수 있다.

도면의 간단한 설명

[31] 도 1은 본 발명의 한 실시예에 따른 동일 사용자 추적 시스템의 구성도이다.

[32] 도 2는 본 발명의 한 실시예에 따른 닉네임 추출 장치가 동일 사용자로 예측되는 닉네임 사용자를 추출하는 방법의 흐름도이다.

[33] 도 3은 본 발명의 한 실시예에 따른 닉네임 추출 장치가 단어 특징 정보를 이용하여 동일 사용자인 것으로 예측되는 닉네임 사용자를 추출한 결과의 정확도를 나타내는 도면이다.

[34] 도 4는 본 발명의 한 실시예에 따른 닉네임 추출 장치가 커뮤니케이션 관계 특징 정보를 이용하여 동일 사용자인 것으로 예측되는 닉네임 사용자를 추출한 결과의 정확도를 나타내는 도면이다.

[35] 도 5 및 도 6은 본 발명의 한 실시예에 따른 닉네임 추출 장치가 특징 정보를 이용하여 동일 사용자를 추론한 예시적 결과이다.

[36] 도 7은 본 발명의 제1 실시예 및 제2 실시예와 종래 기술을 비교한 결과이다.

발명의 실시를 위한 형태

[37] 아래에서는 첨부한 도면을 참고로 하여 본 발명의 실시예에 대하여 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자가 용이하게 실시할 수 있도록 상세히 설명한다. 그러나 본 발명은 여러 가지 상이한 형태로 구현될 수 있으며 여기에서 설명하는 실시예에 한정되지 않는다. 그리고 도면에서 본 발명을 명확하게 설명하기 위해서 설명과 관계없는 부분은 생략하였으며, 명세서 전체를 통하여 유사한 부분에 대해서는 유사한 도면 부호를 붙였다.

[38] 명세서 전체에서, 어떤 부분이 어떤 구성요소를 "포함"한다고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 구성요소를 제외하는 것이 아니라 다른 구성요소를 더 포함할 수 있는 것을 의미한다. 또한, 명세서에 기재된 "...부", "...기", "모듈" 등의 용어는 적어도 하나의 기능이나 동작을 처리하는 단위를 의미하며, 이는 하드웨어나 소프트웨어 또는 하드웨어 및 소프트웨어의 결합으로 구현될 수 있다.

- [39] 다음에서, 온라인 커뮤니티는 주로 특정 취미에 따라 개설된 게시판에 게시글을 올리고 해당 주제에 관심 있는 사용자가 답글을 달거나, 공감을 표시하거나, 다른 커뮤니티로 공유하는 형태로 구성될 수 있다.
- [40] 다음에서, 온라인 커뮤니티 내에서 닉네임을 변경함에 따라 과거의 신원과의 연속성이 단절되는 특성에 따라 닉네임 변경에도 동일한 사용자를 추출할 것을 그 예로 들었으나, 본 발명은 반드시 이에 한정하는 것은 아니며 동일 사용자가 복수의 계정을 생성하고, 각각 생성한 계정으로 소셜 네트워크 서비스를 이용할 때, 동일 사용자를 추출하는 방법으로도 확장 가능하다.
- [41] 또한, 다음에서 설명하는 동일한 사용자를 추출하는 장치 및 방법은 온라인 커뮤니티 내에서 닉네임 변경에 따른 익명성 보장이 어느 정도 수준으로 이루어지는지 평가할 수 있는 방법으로 활용 가능하다.
- [42] 온라인 커뮤니티에서는 익명성이 일정 수준 보장되기 때문에 보다 자유롭게 자신의 의견을 드러낼 수 있다. 온라인 커뮤니티에서 자신이 드러낸 신상 정보에 의해 개인이 식별될 가능성이 있기 때문에, 자신의 신원이 노출되는 것을 꺼리는 사용자들은 닉네임을 변경함으로써 자신의 익명성을 보장하고자 하는데, 닉네임 추적에 따라 동일인으로 식별될 수 있는 정도를 정량적으로 분석함으로써, 익명성 보장에 대한 기준을 마련할 수 있다.
- [43] 또한, 다음에서 설명하는 동일한 사용자를 추출하는 장치 및 방법은 온라인 커뮤니티 내에서 익명성 보장을 악용한 사이버 괴롭힘(cyber bullying)을 차단하는 방법으로 활용 가능하다. 즉, 닉네임 추적을 통해 동일 사용자를 식별하고 시스템적으로 어뷰징을 일삼는 사용자를 차단할 수 있다.
- [44]
- [45] 도 1은 본 발명의 한 실시예에 따른 동일 사용자 추적 시스템의 구성도이다.
- [46] 도 1을 참고하면, 사용자 추적 시스템(10)은 온라인 커뮤니티 제공 서버(200), 그리고 온라인 커뮤니티 제공 서버(200)로부터 온라인 커뮤니티 내에 업로드된 게시글 정보를 전송받아 동일 사용자의 닉네임을 추출하는 닉네임 추출 장치(100)를 포함할 수 있다.
- [47] 온라인 커뮤니티 제공 서버(200)는 온라인 커뮤니티 내에 업로드된 게시글 정보를 닉네임 추출 장치(100)로 제공한다. 온라인 커뮤니티 제공 서버(200)는 일정 기간의 온라인 커뮤니티의 게시글을 수집하고, 게시글에 나타난 정보, 예를 들어 게시글 작성자 닉네임, 게시글 텍스트, 게시글 작성 시각 정보, 게시글 답글 작성자 닉네임, 게시글 답글 텍스트, 게시글 답글 작성 시각 정보, 게시글 공감자 닉네임, 게시글 공유자 닉네임 등의 정보를 데이터베이스화하여 저장하고 있을 수 있다.
- [48] 닉네임 추출 장치(100)는 적어도 하나의 프로세서에 의해 동작하고, 게시글 수신부(110), 특징 정보 추출부(120), 그리고 동일 사용자 추출부(130)를 포함한다.
- [49] 게시글 수신부(110)는 온라인 커뮤니티 제공 서버(200)로부터 온라인

커뮤니티의 게시판에 업로드된 게시글을 수신한다. 본 실시예에서 게시글은 게시판에 작성되는 제목을 포함하여 텍스트, 이미지, 그리고 영상 중 적어도 하나 이상의 내용을 포함하는 일반 글, 일반 글에 대하여 온라인 커뮤니티 사용자의 생각을 짧은 문장, 이모티콘 등으로 표현한 답글을 포함할 수 있다.

[50] 특징 정보 추출부(120)는 게시글 수신부(110)에서 수신한 게시글에서 게시글을 작성한 사용자를 식별할 수 있는 특징 정보를 추출한다. 특징 정보 추출부(120)가 추출하는 특징 정보의 종류는 다양할 수 있다.

[51] 특징 정보 추출부(120)는 해당 커뮤니티 내에 업로드한 게시글에 반영된 사용자의 문체 특징, 게시글에 대한 타 사용자와의 답글, 공감, 공유 정보를 이용한 사용자 간의 커뮤니케이션 관계 특징, 게시글 업로드 시각 정보를 이용한 커뮤니티 활동 시각 특징, 게시글 길이, 사용 단어 개수, 이모티콘 개수, 이미지 업로드 수, 그리고 이미지 정보를 이용한 게시글 특징 등을 추출하여 사용자 고유의 행동 패턴을 식별하도록 할 수 있다.

[52] 한 실시예에 따르면 특징 정보 추출부(120)는 단어 특징을 추출할 수 있다. 온라인 커뮤니티 사용자는 일반글과 답글을 포함하는 게시글을 작성하고, 여기에는 작성자의 관심사나 취향과 같은 사용자 고유의 특징이 반영된다. 특징 정보 추출부(120)는 게시글의 제목, 게시글의 내용, 게시글의 답글을 형태소 분석하여 사용자의 고유 특징을 반영한 특징 단어들을 추출할 수 있다.

[53] 한 실시예에 따르면 특징 정보 추출부(120)는 활동 시각 특징을 추출할 수 있다. 온라인 커뮤니티 사용자는 라이프 스타일에 따라 서로 다른 활동 시각을 가질 수 있다. 예를 들면 직장인들은 점심시간이나 출퇴근 시간, 또는 퇴근시간 이후에 주로 커뮤니티를 사용할 수 있고, 학생은 학기 중과 방학 중에 커뮤니티의 활동 시간 패턴이 달라질 수 있다. 특징 정보 추출부(120)는 일반글, 답글을 포함하는 게시글의 작성 시각 정보를 추출할 수 있다.

[54] 한 실시예에 따르면 특징 정보 추출부(120)는 게시글 특징을 추출할 수 있다. 온라인 커뮤니티 사용자는 자신의 특징에 따라 게시글 사용 패턴이 달라질 수 있다. 예를 들어 온라인 커뮤니티 사용자 중에는 이미지 업로드를 주로 수행하는 사용자, 장문의 글을 게시하는 사용자, 일반글은 게시하지 않고 답글만을 작성하는 사용자 등을 포함할 수 있다. 특징 정보 추출부(120)는 게시글에 포함된 텍스트 내용의 길이, 사용 단어 개수, 이모티콘 개수, 이미지 업로드 수, 그리고 이미지 정보를 추출할 수 있다.

[55] 한 실시예에 따르면 특징 정보 추출부(120)는 커뮤니케이션 관계 특징을 추출할 수 있다. 온라인 커뮤니티에는 소셜 네트워크 서비스에 존재하는 친구 맺기, 또는 팔로워/팔로잉 등으로 명시적으로 드러나는 친구 리스트가 없다. 따라서 일반글과 일반글에 달리는 답글을 통해 간접적으로 사용자들 사이의 관계를 확인할 수 있다. 온라인 커뮤니티 활동 시간이 길어 질수록 답글을 통해 교류하는 사용자들이 형성될 수 있고, 닉네임이 변경된 이후에도 고유하게 유지될 수 있다. 특징 정보 추출부(120)는 사용자가 작성한 게시글에 대하여

답글을 작성하거나, 공감을 표현하거나, 다른 커뮤니티로 게시글을 공유한 다른 닉네임 사용자 목록을 추출하고, 닉네임 사용자마다 답글을 작성하거나, 공감을 표현하거나, 공유한 게시글의 작성자 목록을 추출하여 사용자간의 커뮤니케이션 관계 정보를 추출할 수 있다.

- [56] 동일 사용자 추출부(130)는 특징 정보 추출부(120)에서 추출한 다양한 특징 정보 중 적어도 하나 이상의 특징 정보를 이용하여 복수의 닉네임 집합에서 동일 사용자가 이용한 것으로 예측할 수 있는 닉네임 그룹을 추출한다.
- [57] 한 실시예에 따르면 동일 사용자 추출부(130)는 복수의 닉네임 집합에서 특징 정보를 이용하여 동일 사용자가 이용한 것으로 예측되는 닉네임 그룹을 추출하기 전에, 동일한 사용자가 아닌 것으로 확정되는 닉네임 그룹을 제외할 수 있다.
- [58] 즉, 동일한 사용자라면 일정 사용 기간 이내에서는 두 개 이상의 닉네임을 사용할 수는 없기 때문에 사용 기간이 중첩하는 닉네임 그룹은 동일한 사용자가 아닌 것으로 확정하고, 사용 기간이 중첩하는 닉네임 그룹을 제외하여 동일 사용자인 것으로 예측되는 닉네임 후보 그룹을 추출할 수 있다.
- [59] 동일 사용자 추출부(130)는 어느 하나의 닉네임 사용자와 동일 사용자 인 것으로 예측되는 닉네임 후보 그룹에 속하는 복수의 닉네임 사용자와의 특징 정보 유사도를 각각 계산할 수 있다.
- [60] 동일 사용자 추출부(130)는 특징 정보 추출부(120)에서 추출한 단어 특징 정보, 활동 시각 특징 정보, 게시글 특징, 커뮤니케이션 관계 특징 중 적어도 하나 이상의 특징 정보를 이용하여 특징 정보의 유사 정도가 일정 기준 이상인 닉네임에 대하여 동일 사용자인 것으로 결정할 수 있다.
- [61] 한편, 한 실시예에서 동일 사용자 추출부(130)는 특징 정보 추출부(120)에서 추출한 단어 특징 정보, 활동 시각 특징 정보, 게시글 특징, 커뮤니케이션 관계 특징 정보에 가중치를 설정하고, 가중 합을 계산하여 특징 정보의 유사 정도가 일정 기준 이상인 닉네임에 대하여 동일 사용자인 것으로 결정할 수도 있다.
- [62] 특징 정보의 종류에 따라 어느 특징 정보는 실제 사용자의 고유 특성을 매우 높은 정확도로 반영하고 있는 반면, 일정 특징 정보는 실제 사용자의 고유 특성을 미진한 정확도로 반영할 수 있다. 따라서, 동일 사용자 추출부(130)는 단어 특징 정보, 활동 시각 특징 정보, 게시글 특징 정보, 커뮤니케이션 관계 특징 정보에 서로 다른 파라미터의 가중치를 부여하여 신뢰성이 높은 결과를 획득할 수 있다.
- [63] 동일 사용자 추출부(130)는 유전자 알고리즘(genetic algorithms, GA)를 통해 최적의 파라미터를 추출할 수 있다. 먼저, 동일 사용자 추출부(130)는 각 특징 정보에 대한 가중치를 임의로 설정한 50개의 개체를 무작위로 선택하여 초기 세대를 생성한다. 그리고 각각 단계마다 각 개체들에 대한 적합도를 수행하고, 각 개체에 할당된 가중치를 적용하여 얻은 정확도를 해당 개체의 적합 정도로 판단한다.

- [64] 이후, 각 개체의 후손을 남기기 위하여 선별, 교배, 돌연변이 방법을 선택하는데, 선별은 적합 정도가 높은 상위 10개의 개체를 다음 세대로 전달하는 것을 의미하고, 교배는 상위 10개의 개체를 제외한 나머지 개체 중 30개의 개체를 두 개씩 짝지어 가중치를 반반 섞는 것을 의미한다. 그리고 나머지 선별, 교배가 이루어지지 않은 나머지 개체 10개는 임의의 특징 정보를 무작위로 변경한다.
- [65] 상술한 과정을 수회, 수십 회 반복한 다음, 적합도가 가장 높은 개체를 특징 정보에 대한 최적의 파라미터로 설정할 수 있다.
- [66] 이와 같이 본 발명의 실시예에 따른 닉네임 추출 장치(100)는 온라인 커뮤니티에서 사용자가 닉네임을 변경하더라도 동일 사용자의 닉네임 목록을 추출함으로써 서로 다른 닉네임으로 활동한 경우에도 동일 인물에 대한 수집 데이터 단절을 방지할 수 있다.
- [67]
- [68] 도 2는 본 발명의 한 실시예에 따른 닉네임 추출 장치가 동일 사용자로 예측되는 닉네임 사용자를 추출하는 방법의 흐름도이다.
- [69] 닉네임 추출 장치(100)는 온라인 커뮤니티 제공 서버(200)로부터 온라인 커뮤니티 내에 업로드 된 게시글 정보를 수신한다(S110).
- [70] 닉네임 추출 장치(100)는 게시글 정보에서 게시글을 작성한 사용자를 식별하는 특징 정보를 추출한다(S120). 닉네임 추출 장치(100)가 추출하는 특징 정보의 종류는 다양할 수 있다. 예를 들어, 닉네임 추출 장치(100)는 작성자의 관심사나 취향과 같은 사용자 고유의 특징이 반영된 단어 특징 정보, 게시글의 작성 시각 정보, 게시글 특징 정보, 타 사용자들과의 커뮤니케이션 관계 특징 정보를 추출할 수 있다.
- [71] 단어 특징 정보는 게시글의 제목, 게시글의 내용, 게시글의 답글을 형태소 분석을 통해 추출될 수 있다. 게시글 특징 정보는 게시글에 포함된 텍스트 내용의 길이, 사용 단어 개수, 이모티콘 개수, 이미지 업로드 수, 그리고 이미지 정보의 평균 및 분산 값을 이용하여 추출될 수 있다. 커뮤니케이션 관계 특징 정보는 사용자가 작성한 게시글에 대하여 답글을 작성하거나, 공감을 표현하거나, 다른 커뮤니티로 게시글을 공유한 다른 닉네임 사용자 목록을 추출하고, 닉네임 사용자마다 답글을 작성하거나, 공감을 표현하거나, 공유한 게시글의 작성자 목록을 이용하여 추출될 수 있다.
- [72] 닉네임 추출 장치(100)는 복수의 닉네임 집합에서 동일한 사용자가 아닌 것으로 확정되는 닉네임 그룹을 제외하여 후보 닉네임 집합을 선별한다(S130).
- [73] 동일한 사용자라면 온라인 커뮤니티 내의 일정 사용 기간 이내에서는 두 개 이상의 닉네임을 사용할 수는 없기 때문에 사용 기간이 중첩하는 닉네임은 동일한 사용자가 아닌 것으로 확정하고, 사용 기간이 중첩하는 닉네임을 제외하여 후보 닉네임 집합을 선별할 수 있다.
- [74] 전체 닉네임 집합 $N = \{N_1, N_2, \dots, N_m\}$ 중 어느 하나의 원소인 N_i 에 대하여 사용

인터벌 $NT_{Ni}=[t_p, t_q]$ 을 가질 때, 후보 닉네임 집합(CS_{Ni})은 수학식 1과 같이 정의될 수 있다.

[75] [수학식 1]

$$[76] \quad CS_{N_i} = \{ N_j \in N \mid NT_{N_i} \cap NT_{N_j} = \emptyset \}$$

[77] 그리고 닉네임 추출 장치(100)는 추출한 특징 정보를 기초로 후보 닉네임 집합에서 동일한 사용자가 이용한 것으로 예측되는 닉네임 사용자를 추출한다(S140).

[78] 닉네임 추출 장치(100)는 어느 하나의 닉네임 사용자의 특징 정보, 그리고 닉네임 사용자와 동일한 사용자일 것으로 예측되는 후보 닉네임 집합에 포함된 각 닉네임 사용자의 특징 정보와의 유사 정도를 계산한다.

[79] 본 실시예에서 닉네임 추출 장치(100)는 하기 표 1과 같이 단어 특징 정보 및 커뮤니케이션 관계 특징 정보에 대해서는 복수의 닉네임 사용자와의 특징 정보와의 자카드 유사도를 이용하여 유사 정도를 계산할 수 있고, 활동 시각 특징 정보에 대해서는 동적 타임 워핑(Dynamic time warping) 알고리즘을 이용하여 유사 정도를 계산할 수 있으며, 게시글 특징 정보에 대해서는 평균 및 분산 값을 이용하여 유사 정도를 계산할 수 있다.

[80] [Table 1]

특징 정보 종류	세부 특징 정보	metric
단어 특징 정보	일반글 제목 특징 단어	자카드 유사도
	일반글 내용 특징 단어	자카드 유사도
	답글 특징 단어	자카드 유사도
활동 시각 특징 정보	일반글 작성 시간	동적타임워핑(DTW)
	답글 작성 시간	동적타임워핑(DTW)
게시글 특징 정보	일반글/답글 작성 비율	비율
	일반글 제목 길이	평균
		분산
	일반글 내용 길이	평균
		분산
	일반글 제목 단어 수	평균
		분산
	일반글 내용 단어 수	평균
		분산
	답글 수	평균
		분산
	이미지 수	평균
		분산
	조회수	평균
		분산
커뮤니케이션 관계 특징 정보	게시글에 대해 반응을 보인 타 사용자 목록(제1목록)	자카드 유사도
	반응을 보인 타 게시글의 작성자 목록(제2 목록)	자카드 유사도
	합집합(제1 목록, 제2목록)	자카드 유사도
	교집합(제1 목록, 제2 목록)	자카드 유사도

[81] 그리고 닉네임 추출 장치(100)는 후보 닉네임 집합에 포함된 닉네임 사용자

중에서 특징 정보의 유사 정도가 일정 기준 이상인 닉네임을 추출하여 동일 사용자인 것으로 결정할 수 있다.

[82] 표 1과 같이 단어 특징 정보, 활동 시각 특징 정보, 게시글 특징 정보, 그리고 커뮤니케이션 관계 특징 정보에 포함된 세부 특징 정보(총 24개) $F=\{f_1, f_2, \dots, f_{24}\}$ 에 대해 N_i 와 CS_{N_i} 에 속하는 후보 닉네임들과의 유사도를 키 값(key-value) 형태로 구하고 이를 원소로 하는 집합의 형태로 리턴 해주는 함수를 $\text{sim}_{f_i}(CS_{N_i})$ 라고 정의한다.

[83] 그리고 이때 어느 하나의 세부 특징 정보 f_j 에 대한 후보군 집합은 다음의 수학적 식 2와 같이 정의할 수 있다.

[84] [수학적 식 2]

[85] $CS_{N_i, f_j} = \text{sim}_{f_j}(CS_{N_i})$

[86] 그리고 후보군 집합에 속하는 닉네임 중에서 유사도가 높은 상위 k 개 닉네임들을 원소로 가지는 집합을 최종 후보군 집합 SubCS_{N_i, f_j} 으로 정의할 수 있다.

[87] 한편, 다른 실시예에서 동일 사용자 추출부(130)는 특징 정보 추출부(120)에서 추출한 단어 특징 정보, 활동 시각 특징 정보, 게시글 특징, 커뮤니케이션 관계 특징 정보에 가중치를 설정하고, 가중 합을 계산하여 특징 정보의 유사 정도가 일정 기준 이상인 닉네임에 대하여 동일 사용자인 것으로 결정할 수도 있다. 이때, 동일 사용자 추출부(130)는 유전자 알고리즘(genetic algorithms, GA)를 통해 최적의 가중치를 추출하여 신뢰성이 높은 결과를 획득할 수 있다.

[88]

[89] 도 3은 본 발명의 한 실시예에 따른 닉네임 추출 장치가 단어 특징 정보를 이용하여 동일 사용자인 것으로 예측되는 닉네임 사용자를 추출한 결과의 정확도를 나타내는 도면이고, 도 4는 본 발명의 한 실시예에 따른 닉네임 추출 장치가 커뮤니케이션 관계 특징 정보를 이용하여 동일 사용자인 것으로 예측되는 닉네임 사용자를 추출한 결과의 정확도를 나타내는 도면이다.

[90] 한편, 본 발명의 한 실시예에서 N_i 에 대한 동일 사용자 추론의 정확도는 하기의 수학적 식 3과 같이 구하였다.

[91] [수학적 식 3]

[92]
$$Acc_{N_i} = \frac{\text{SubCS}_{N_i, f_j} \cap \text{Sol}_{N_i}}{\text{SubCS}_{N_i, f_j}}$$

[93] 여기서, Sol_{N_i} 는 실제 동일 사용자가 사용한 닉네임들의 집합이다.

[94] 도 3을 참고하면, 일반글의 제목에 포함된 단어 특징 정보, 일반글의 내용에 포함된 단어 특징 정보, 답글에 포함된 단어 특징 정보 순서대로 동일 사용자 추론에 대해 높은 정확도를 보이는 것을 확인할 수 있다.

[95] 전반적으로 답글보다는 일반글과 관련된 단어들이 높은 정확도를 보이는데, 이는 일반적으로 일반글의 제목이나 내용은 사용자의 관심사나 취향을

반영하여 개인별로 뚜렷한 특징을 보일 수 있는 단어들이 많이 사용되기 때문이다.

[96] 반면에 답글은 해당 답글이 작성된 일반글의 내용에 대한 의견 표현이 주를 이루기 때문에 상대적으로 낮은 정확도인 것을 확인할 수 있다.

[97]

[98] 도 4를 참고하면, 특정 닉네임 사용자가 작성한 게시글에 대하여 답글을 작성하거나, 공감을 표현하거나, 공유한 다른 사용자(제1 목록) 특징 정보가 특정 닉네임 사용자가 반응을 보인 타 게시글의 작성자(제2 목록) 특징 정보보다 높은 정확도를 갖는다. 이는 제1 목록 특징 정보가 제2 목록 특징 정보에 비하여 누적된 데이터를 포함할 수 있기 때문에 특정 닉네임을 식별하는데에 더욱 높은 정확도를 가지는 것으로 해석할 수 있다.

[99] 한편, 제1 목록과 제2 목록의 합집합으로부터 도출한 특징 정보가 제1 목록과 제2 목록의 교집합으로부터 도출한 특징 정보보다 높은 정확도를 갖는다. 마찬가지로 합집합으로부터 도출한 특징 정보는 교집합으로부터 도출한 특징 정보에 비하여 누적된 데이터를 다량 포함할 수 있으므로, 특정 닉네임을 식별하는데에 더욱 높은 정확도를 가지는 것으로 해석할 수 있다.

[100]

[101] 도 5 및 도 6은 본 발명의 한 실시예에 따른 닉네임 추출 장치가 특징 정보를 이용하여 동일 사용자를 추론한 예시적 결과이다.

[102] 도 5는 닉네임에 대한 특징 정보별 정확도를 시각화한 hitmap이며, 열은 닉네임이고, 행은 특징 정보들이다. 각 셀은 정답률을 의미하는 것으로 어둡게 나타나는 셀은 특징 정보에 대한 정답률이 높은 것을 의미한다. 대체적으로 단어 특징 정보와 커뮤니케이션 관계 특징 정보가 높은 정확도를 보이는 것을 확인할 수 있었다.

[103] 도 6은 도 5의 결과를 정밀하게 분석하기 위하여 위해 X-means 알고리즘을 통해 클러스터링한 결과이다.

[104] 클러스터 1의 경우, 특징 정보가 닉네임 식별 정보를 충분히 반영하지 않아 닉네임 식별을 할 수 없는 결과를 나타낸다.

[105] 클러스터 2는 단어 특징 정보들에서 높은 정답률을 갖는 결과를 확인할 수 있다.

[106] 클러스터 3, 4는 단어 특징 정보, 커뮤니케이션 관계 특징 정보들에서 높은 정답률을 갖는 결과를 확인하였다.

[107] 도 3 내지 도 6의 결과를 참고하면, 동일 사용자 추론의 정확도는 각 특징 정보마다 다른 정확도를 나타내는 것을 확인할 수 있다.

[108] 따라서, 본 발명의 다른 실시예에 따른 닉네임 추출 장치(100)는 각 특징 정보에 서로 다른 가중치를 부여하여 특징 정보들의 가중 합을 통해 동일 사용자를 추론할 수 있다.

[109] 도 7은 본 발명의 제1 실시예 및 제2 실시예와 종래 기술을 비교한 결과이다.

- [110] 도 7을 참고하면, 종래 기술(jaro distance)에 따라 특징 정보의 유사도를 계산하여 동일 사용자를 추론한 결과의 정확도는 약 17%로 비교적 낮은 것을 확인할 수 있다.
- [111] 반면 본 발명의 제1 실시예에 따른 닉네임 추출 장치가 게시글에서 추출한 복수의 특징 정보를 이용하여 동일 가중치를 두고, 동일 사용자를 추론한 결과의 정확도는 약 55%인 것으로 종래 기술에 비해 정확도가 향상된 것을 확인할 수 있다.
- [112] 한편, 본 발명의 제2 실시예에 따른 닉네임 추출 장치가 게시글에서 추출한 복수의 특징 정보에 서로 다른 가중치를 부여하여 가중 합을 통해 동일 사용자를 추론한 결과는 정확도가 약 75%로 종래 기술 및 제1 실시예에 비하여 더욱 향상한 것을 확인할 수 있다.
- [113] 이와 같이, 본 발명의 실시예에 따르면 닉네임 변경이 가능한 온라인 커뮤니티 내에서 동일 사용자에 속하는 닉네임을 추출함으로써, 익명성이 보장된 온라인 커뮤니티에서도 개인에 대한 프로파일링을 수행할 수 있다. 이러한 프로파일링 결과를 이용하여 기업의 의사 결정이나 사회 현상을 분석하기 적합한 정보를 획득할 수 있다.
- [114] 이상에서 설명한 본 발명의 실시예는 장치 및 방법을 통해서만 구현이 되는 것은 아니며, 본 발명의 실시예의 구성에 대응하는 기능을 실현하는 프로그램 또는 그 프로그램이 기록된 기록 매체를 통해 구현될 수도 있다.
- [115] 이상에서 본 발명의 실시예에 대하여 상세하게 설명하였지만 본 발명의 권리범위는 이에 한정되는 것은 아니고 다음의 청구범위에서 정의하고 있는 본 발명의 기본 개념을 이용한 당업자의 여러 변형 및 개량 형태 또한 본 발명의 권리범위에 속하는 것이다.

청구범위

- [청구항 1] 적어도 하나의 프로세서에 의해 동작하는 장치가 동일 사용자의 닉네임을 추출하는 방법으로서,
서버로부터 온라인 커뮤니티 내에 업로드 된 게시글을 수신하는 단계,
상기 게시글에서 상기 게시글을 작성한 사용자를 식별하는 적어도 하나 이상의 특징 정보를 추출하는 단계, 그리고
복수의 닉네임 집합에서 상기 특징 정보의 유사 정도가 일정 기준 이상인 닉네임을 동일 사용자의 닉네임으로 추출하는 단계를 포함하는 닉네임 추출 방법.
- [청구항 2] 제1항에서,
상기 복수의 닉네임 집합에서 동일 사용자가 아닌 것으로 확정되는 닉네임을 제외하여 후보 닉네임 집합을 선별하는 단계를 더 포함하고,
상기 닉네임 추출 단계는
상기 후보 닉네임 집합에서 상기 특징 정보의 유사 정도가 일정 기준 이상인 닉네임을 동일 사용자의 닉네임으로 추출하는 닉네임 추출 방법.
- [청구항 3] 제2항에서,
상기 동일 사용자가 아닌 것으로 확정되는 닉네임은 사용 기간이 중첩되는 닉네임인 닉네임 추출 방법.
- [청구항 4] 제1항에서,
상기 특징 정보를 추출하는 단계는
상기 사용자의 관심 정보를 식별할 수 있는 단어 특징 정보, 상기 사용자의 활동 시간 특징 정보, 상기 게시글 특징 정보, 그리고
상기 사용자와 다른 사용자와의 커뮤니케이션 관계 특징 정보 중 적어도 하나 이상의 특징 정보를 추출하는 닉네임 추출 방법.
- [청구항 5] 제4항에서,
상기 단어 특징 정보는
상기 게시글 텍스트에서 형태소 분석을 이용하여 적어도 하나 이상의 특징 단어, 그리고 상기 특징 단어의 사용 빈도 수 정보를 포함하는 닉네임 추출 방법.
- [청구항 6] 제4항에서,
상기 커뮤니케이션 관계 특징 정보는
상기 사용자가 작성한 적어도 하나의 게시글에 대하여 답글을 작성하거나 공감을 표현한 닉네임 정보인 제1 닉네임 특징 정보,
상기 사용자가 답글을 작성하거나, 공감을 표현한 타 사용자의

[청구항 7]

게시글 작성자 닉네임 정보인 제2 닉네임 특징 정보, 그리고
상기 제1 닉네임 특징 정보와 상기 제2 닉네임 특징 정보를 조합한
특징 정보 중 적어도 하나 이상인 닉네임 추출 방법.

적어도 하나의 프로세서에 의해 동작하는 장치가 동일 사용자의
닉네임을 추출하는 방법으로서,
서버로부터 온라인 커뮤니티 내에 업로드 된 게시글을 수신하는
단계,

상기 게시글에서 사용자의 관심 정보를 식별할 수 있는 단어 특징
정보, 상기 사용자의 활동 시간 특징 정보, 상기 게시글 특징 정보,
그리고 상기 사용자와 다른 사용자와의 커뮤니케이션 관계 특징
정보 중 적어도 둘 이상을 포함하는 특징 정보를 추출하는 단계,
복수의 닉네임 간 상기 특징 정보들의 유사 정도를 계산하는 단계,
상기 특징 정보마다 서로 다른 가중값을 상기 유사 정도에
부여하는 단계, 그리고

상기 유사 정도가 일정 기준 이상인 닉네임을 동일 사용자의
닉네임으로 추출하는 단계를 포함하는 닉네임 추출 방법.

[청구항 8]

제7항에서,

상기 복수의 닉네임에서 동일 사용자가 아닌 것으로 확정되는
닉네임을 제외하여 후보 닉네임 집합을 선별하는 단계를 더
포함하고,

상기 닉네임 추출 단계는

상기 후보 닉네임 집합에서 상기 유사 정도가 일정 기준 이상인
닉네임을 동일 사용자의 닉네임으로 추출하는 닉네임 추출 방법.

[청구항 9]

제8항에서,

상기 동일 사용자가 아닌 것으로 확정되는 닉네임은 사용 기간이
중첩되는 닉네임인 닉네임 추출 방법.

[청구항 10]

제7항에서,

상기 가중값은 유전 알고리즘을 통해 산출된 값인 닉네임 추출
방법.

[청구항 11]

동일한 사용자의 닉네임을 추출하기 위한 장치로서,

온라인 커뮤니티에 업로드된 적어도 하나의 게시글을 수신하는
게시글 수신부,

상기 게시글에서 상기 게시글을 작성한 사용자를 식별하는 적어도
하나 이상의 특징 정보를 추출하는 특징 정보 추출부, 그리고
상기 특징 정보를 기초로 복수의 닉네임 집합에서 상기 특징
정보의 유사 정도가 일정 기준 이상인 닉네임을 동일 사용자의
닉네임으로 추출하는 동일 사용자 추출부를 포함하는 닉네임 추출
장치.

- [청구항 12] 제11항에서,
 상기 동일 사용자 추출부는
 상기 복수의 닉네임 집합에서 동일 사용자가 아닌 것으로
 확정되는 닉네임을 제외하여 후보 닉네임 집합을 선별하고,
 상기 후보 닉네임 집합에서 상기 특징 정보의 유사 정도가 일정
 기준 이상인 닉네임을 동일 사용자의 닉네임으로 추출하는 닉네임
 추출 장치.
- [청구항 13] 제12항에서,
 상기 동일 사용자가 아닌 것으로 확정되는 닉네임은 사용 기간이
 중첩되는 닉네임인 닉네임 추출 장치.
- [청구항 14] 제11항에서,
 상기 특징 정보 추출부는
 상기 사용자의 관심 정보를 식별할 수 있는 단어 특징 정보, 상기
 사용자의 활동 시간 특징 정보, 상기 게시글 특징 정보, 그리고
 상기 사용자와 다른 사용자와의 커뮤니케이션 관계 특징 정보 중
 적어도 하나 이상의 특징 정보를 추출하는 닉네임 추출 장치.
- [청구항 15] 제14항에서,
 상기 단어 특징 정보는
 상기 게시글 텍스트에서 형태소 분석을 이용하여 적어도 하나
 이상의 특징 단어, 그리고 상기 특징 단어의 사용 빈도 수 정보를
 포함하는 닉네임 추출 장치.
- [청구항 16] 제14항에서,
 상기 커뮤니케이션 관계 특징 정보는
 상기 사용자가 작성한 적어도 하나의 게시글에 대하여 답글을
 작성하거나 공감을 표현한 닉네임 정보인 제1 닉네임 특징 정보,
 상기 사용자가 답글을 작성하거나, 공감을 표현한 타 사용자의
 게시글 작성자 닉네임 정보인 제2 닉네임 특징 정보, 그리고
 상기 제1 닉네임 특징 정보와 상기 제2 닉네임 특징 정보를 조합한
 특징 정보 중 적어도 하나 이상인 닉네임 추출 장치.
- [청구항 17] 동일한 사용자의 닉네임을 추출하기 위한 장치로서,
 온라인 커뮤니티에 업로드된 적어도 하나의 게시글을 수신하는
 게시글 수신부,
 상기 게시글에서 사용자의 관심 정보를 식별할 수 있는 단어 특징
 정보, 상기 사용자의 활동 시간 특징 정보, 상기 게시글 특징 정보,
 그리고 상기 사용자와 다른 사용자와의 커뮤니케이션 관계 특징
 정보 중 적어도 둘 이상을 포함하는 특징 정보를 추출하는 특징
 정보 추출부, 그리고
 복수의 닉네임 간 상기 특징 정보들의 유사 정도를 계산하고, 상기

특징 정보마다 서로 다른 가중값을 상기 유사 정도에 부여하며,
상기 유사 정도가 일정 기준 이상인 닉네임을 동일 사용자의
닉네임으로 추출하는 동일 사용자 추출부를 포함하는 닉네임 추출
장치.

[청구항 18]

제17항에서,
상기 동일 사용자 추출부는
상기 복수의 닉네임에서 동일 사용자가 아닌 것으로 확정되는
닉네임을 제외하여 후보 닉네임 집합을 선별하고, 상기 후보
닉네임 집합에서 상기 유사 정도가 일정 기준 이상인 닉네임을
동일 사용자의 닉네임으로 추출하는 닉네임 추출 장치.

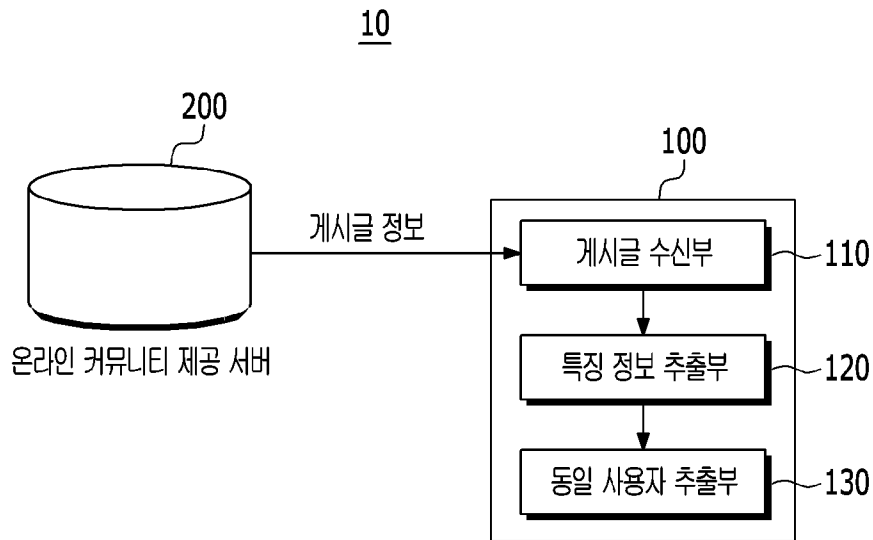
[청구항 19]

제18항에서,
상기 동일 사용자가 아닌 것으로 확정되는 닉네임은 사용 기간이
중첩되는 닉네임인 닉네임 추출 장치.

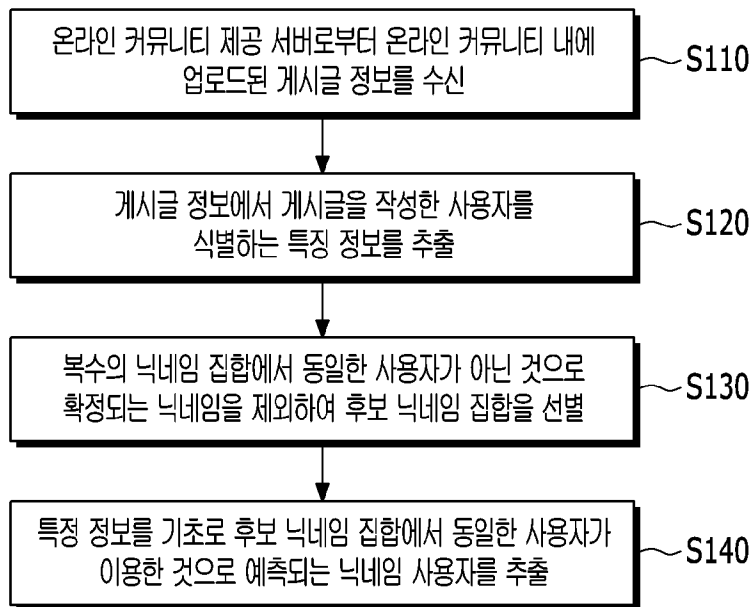
[청구항 20]

제17항에서,
상기 가중값은 유전 알고리즘을 통해 산출된 값인 닉네임 추출
장치.

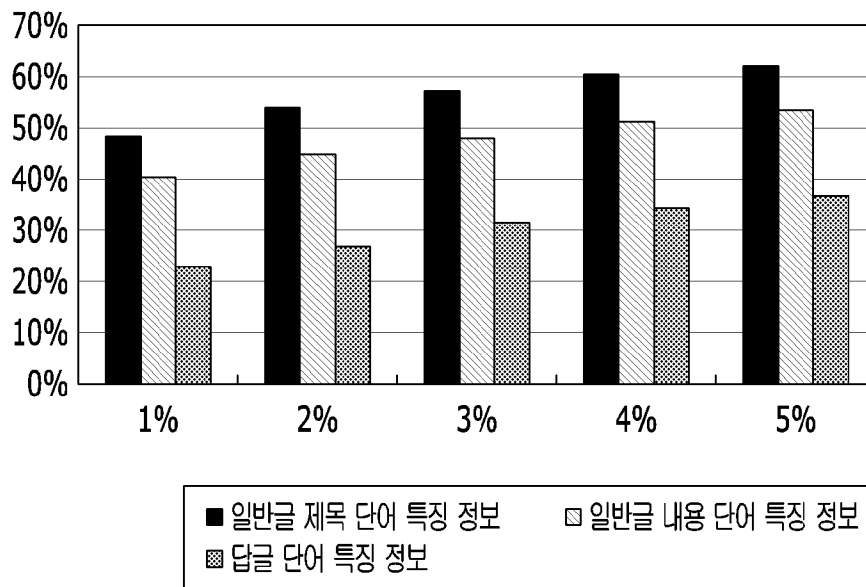
[Fig. 1]



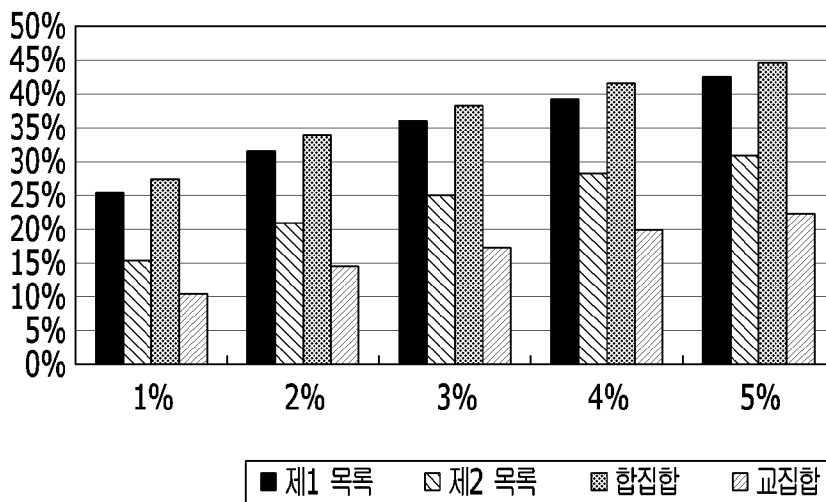
[Fig. 2]



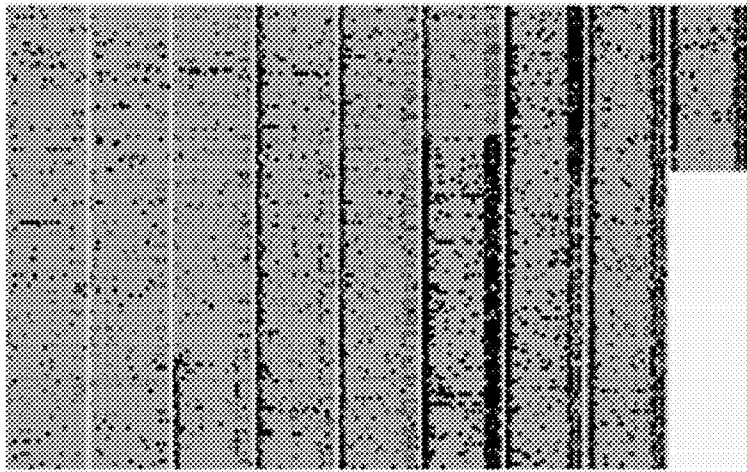
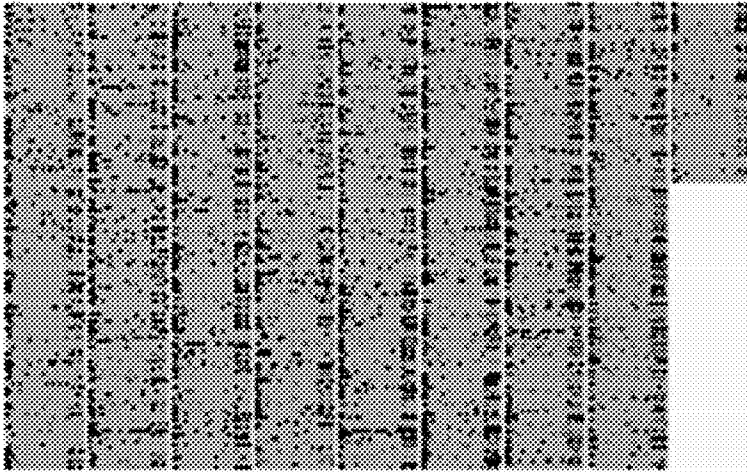
[Fig. 3]



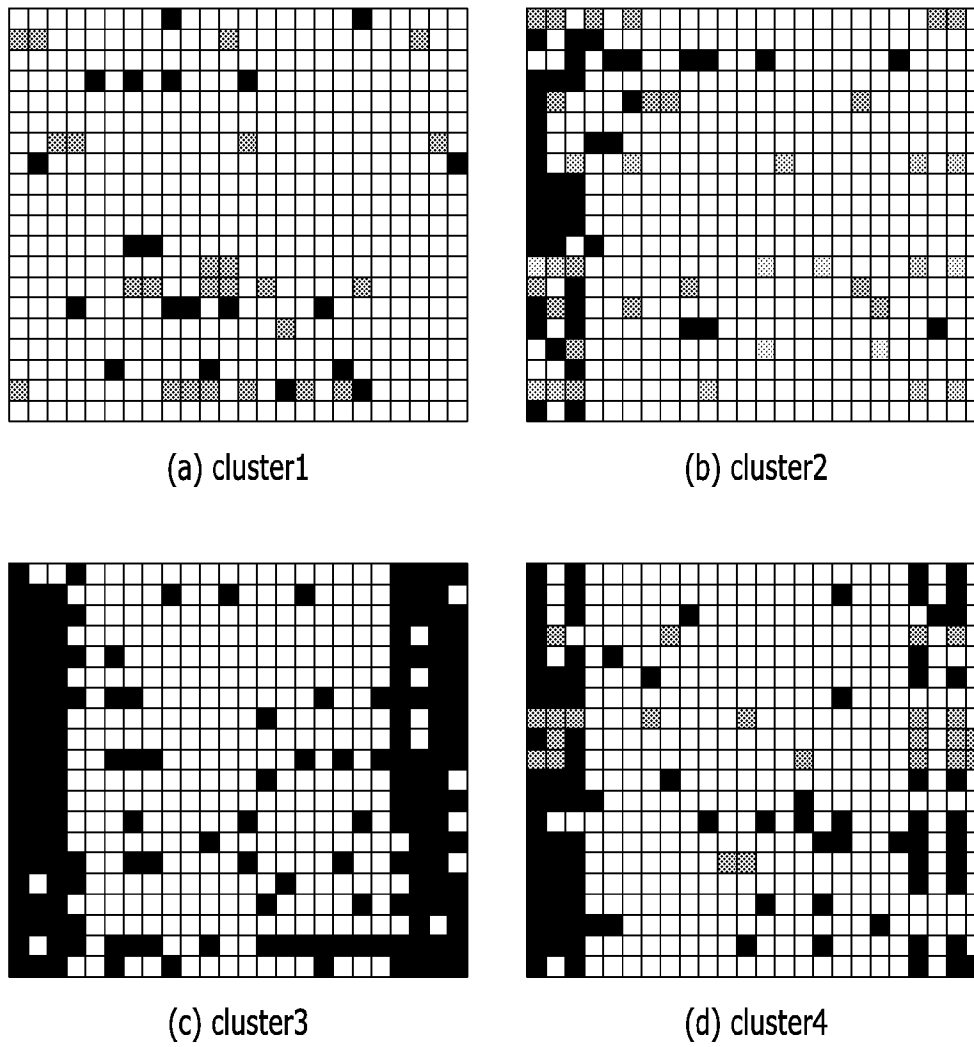
[Fig. 4]



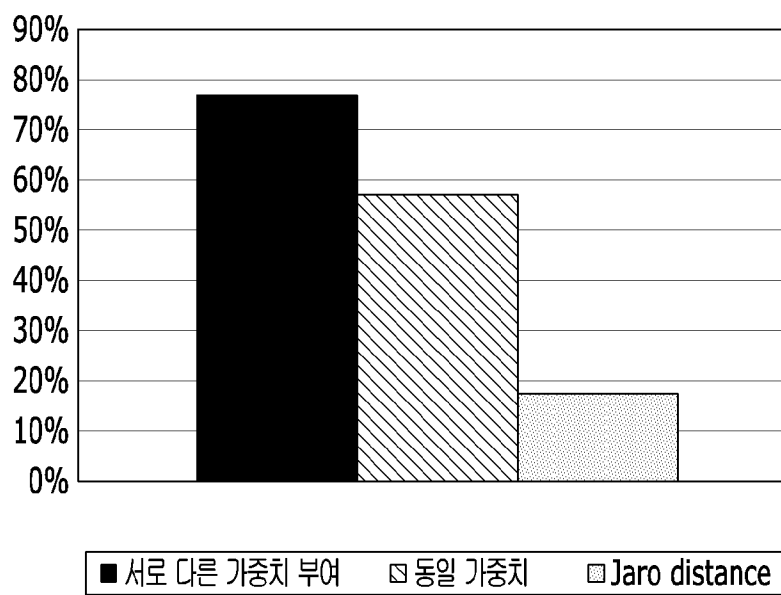
[Fig. 5]



[Fig. 6]



[Fig. 7]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/KR2017/014108**A. CLASSIFICATION OF SUBJECT MATTER****G06Q 50/30(2012.01)i, G06F 17/26(2006.01)i, G06F 17/00(2006.01)i**

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06Q 50/30; G06Q 50/10; G06F 17/20; G06F 17/30; G06Q 30/06; G06Q 50/00; G06F 17/26; G06F 17/00

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Korean Utility models and applications for Utility models: IPC as above

Japanese Utility models and applications for Utility models: IPC as above

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

eKOMPASS (KIPO internal) & Keywords: online, community, posting, characteristic information, nickname, same user, identification

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	KIM, Min Jae et al., "Measures of Abnormal User Activities in Online Comments Based on Cosine Similarity", Journal of the Korea Institute of Information Security & Cryptology, April 2014, vol. 24, no. 2, pages 335-343. See pages 337-341.	1-20
Y	KR 10-2015-0031341 A (TENCENT TECHNOLOGY(SHENZHEN) COMPANY LIMITED) 23 March 2015 See paragraphs [0043]-[0078], claims 1-4 and figures 2-4.	1-20
Y	KR 10-2010-0001650 A (SOL SPECTRUM CO., LTD.) 06 January 2010 See claims 1-3, 7 and figure 1.	6,16
Y	KR 10-2013-0091392 A (FOUNDATION OF SOONGSIL UNIVERSITY-INDUSTRY COOPERATION) 19 August 2013 See paragraph [0060], claims 1, 9 and figure 1.	10,20
A	KR 10-2015-0101536 A (SK PLANET CO., LTD.) 04 September 2015 See claims 1-2, 13 and figures 1, 4.	1-20



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

14 MARCH 2018 (14.03.2018)

Date of mailing of the international search report

14 MARCH 2018 (14.03.2018)

Name and mailing address of the ISA/KR

Korean Intellectual Property Office
Government Complex-Daejeon, 189 Sconsa-ro, Daejeon 302-701,
Republic of Korea

Facsimile No. +82-42-481-8578

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/KR2017/014108

Patent document cited in search report	Publication date	Patent family member	Publication date
KR 10-2015-0031341 A	23/03/2015	CN 103593799 A	19/02/2014
		CN 103593799 B	26/10/2016
		KR 10-1555810 B1	24/09/2015
		TW 201409252 A	01/03/2014
		TW 1516949 B	11/01/2016
		US 2015-0234913 A1	20/08/2015
		WO 2014-026526 A1	20/02/2014
KR 10-2010-0001650 A	06/01/2010	KR 10-2010-0003184 A	07/01/2010
KR 10-2013-0091392 A	19/08/2013	KR 10-1382069 B1	14/04/2014
KR 10-2015-0101536 A	04/09/2015	NONE	

A. 발명이 속하는 기술분류(국제특허분류(IPC))

G06Q 50/30(2012.01)i, G06F 17/26(2006.01)i, G06F 17/00(2006.01)i

B. 조사된 분야

조사된 최소문헌(국제특허분류를 기재)

G06Q 50/30; G06Q 50/10; G06F 17/20; G06F 17/30; G06Q 30/06; G06Q 50/00; G06F 17/26; G06F 17/00

조사된 기술분야에 속하는 최소문헌 이외의 문헌

한국등록실용신안공보 및 한국공개실용신안공보: 조사된 최소문헌란에 기재된 IPC

일본등록실용신안공보 및 일본공개실용신안공보: 조사된 최소문헌란에 기재된 IPC

국제조사에 이용된 전산 데이터베이스(데이터베이스의 명칭 및 검색어(해당하는 경우))

eKOMPASS(특허청 내부 검색시스템) & 키워드: 온라인, 커뮤니티, 게시물, 특정정보, 닉네임, 동일사용자, 식별

C. 관련 문헌

카테고리*	인용문헌명 및 관련 구절(해당하는 경우)의 기재	관련 청구항
Y	김민재 등, ‘코사인 유사도 기반의 인터넷 댓글 상 이상 행위 분석 방법’, 정보보호학회논문지, 2014.04., 제24권, 제2호, 페이지 335-343. 페이지 337-341 참조.	1-20
Y	KR 10-2015-0031341 A (텐센트 테크놀로지(셴젠) 컴퍼니 리미티드) 2015.03.23 단락 [0043]-[0078], 청구항 1-4 및 도면 2-4 참조.	1-20
Y	KR 10-2010-0001650 A (솔스펙트럼(주)) 2010.01.06 청구항 1-3,7 및 도면 1 참조.	6,16
Y	KR 10-2013-0091392 A (승실대학교산학협력단) 2013.08.19 단락 [0060], 청구항 1,9 및 도면 1 참조.	10,20
A	KR 10-2015-0101536 A (에스케이플래닛 주식회사) 2015.09.04 청구항 1-2,13 및 도면 1,4 참조.	1-20

☐ 추가 문헌이 C(계속)에 기재되어 있습니다.

☒ 대응특허에 관한 별지를 참조하십시오.

* 인용된 문헌의 특별 카테고리:

“A” 특별히 관련이 없는 것으로 보이는 일반적인 기술수준을 정의한 문헌

“T” 국제출원일 또는 우선일 후에 공개된 문헌으로, 출원과 상충하지 않으며 발명의 기초가 되는 원리나 이론을 이해하기 위해 인용된 문헌

“E” 국제출원일보다 빠른 출원일 또는 우선일을 가지나 국제출원일 이후에 공개된 선출원 또는 특허 문헌

“X” 특별한 관련이 있는 문헌. 해당 문헌 하나만으로 청구된 발명의 신규성 또는 진보성이 없는 것으로 본다.

“L” 우선권 주장에 의문을 제기하는 문헌 또는 다른 인용문헌의 공개일 또는 다른 특별한 이유(이유를 명시)를 밝히기 위하여 인용된 문헌

“Y” 특별한 관련이 있는 문헌. 해당 문헌이 하나 이상의 다른 문헌과 조합하는 경우로 그 조합이 당업자에게 자명한 경우 청구된 발명은 진보성이 없는 것으로 본다.

“O” 구두 개시, 사용, 전시 또는 기타 수단을 언급하고 있는 문헌

“&” 동일한 대응특허문헌에 속하는 문헌

“P” 우선일 이후에 공개되었으나 국제출원일 이전에 공개된 문헌

국제조사의 실제 완료일

2018년 03월 14일 (14.03.2018)

국제조사보고서 발송일

2018년 03월 14일 (14.03.2018)

ISA/KR의 명칭 및 우편주소



대한민국 특허청
(35208) 대전광역시 서구 청사로 189,
4동 (둔산동, 정부대전청사)

팩스 번호 +82-42-481-8578

심사관

김연정

전화번호 +82-42-481-3325



국제조사보고서에서 인용된 특허문헌	공개일	대응특허문헌	공개일
KR 10-2015-0031341 A	2015/03/23	CN 103593799 A CN 103593799 B KR 10-1555810 B1 TW 201409252 A TW I516949 B US 2015-0234913 A1 WO 2014-026526 A1	2014/02/19 2016/10/26 2015/09/24 2014/03/01 2016/01/11 2015/08/20 2014/02/20
KR 10-2010-0001650 A	2010/01/06	KR 10-2010-0003184 A	2010/01/07
KR 10-2013-0091392 A	2013/08/19	KR 10-1382069 B1	2014/04/14
KR 10-2015-0101536 A	2015/09/04	없음	