

명세서

발명의 명칭: 개인 사전 구축 서비스 제공 시스템 및 방법

기술분야

- [1] 본 발명은 사용자별로 개인화된 문자열-태그 매칭을 통해 개인 사전을 구축하는 서비스를 제공하는 기술에 관한 것이다.

배경기술

- [2] 사람들이 글을 쓸 때, 그 사람의 문학적 자질에 따라 그 정도의 차이는 있겠으나, 정확한 의사를 상대방에게 전달하기 위하여 적절한 문구와 문장 구사 때문에 누구나 고민한적이 있을 것이다.
- [3] 특히, 과거에 작성하였거나 봤었던 좋은 문구들을 다시 기억해내거나 찾기 어려워, 결국 새롭게 글을 다시 작성해야 하는 번거로움과 함께 기존의 문구에 대한 좋은 기억과 달리 새롭게 작성한 글이 마음에 들지 않아 고민했던 경험들을 누구나 가지고 있을 것이다.
- [4] 또한, 소규모 팀단위로 협업을 통해 문서를 작성함에 있어 각 개인 간의 표현이 서로 달라 작업의 통일성 및 효율성을 도모하기 어려웠던 경험도 있을 것이다.
- [5] 최근에는 인공지능 알고리즘의 발달로 단순한 주제어 입력만으로 원하는 문장을 쉽고 편리하게 작성하는 서비스가 출시되고 있으나, 이러한 서비스는 빅데이터에 기반한 기계학습의 결과물로서 범용적인 문구와 문장을 제공할 뿐, 사용자별로 개인화된 문구와 문장을 제공하지는 못하는 문제가 있다.
- [6] 또한, 이와 같은 빅데이터에 기반한 문서 작성 서비스는 보안 문제, 저작권 문제 등에도 노출될 수 밖에 없는 단점이 있다.
- [7] 따라서, 보안 문제, 저작권 문제 등을 해결하고 사용자별 경험에 기반한 문서를 보다 편리하고 효율적으로 작성할 수 있는 문서 작성 서비스에 대한 필요성이 대두되고 있다.
- [8] (특허문헌 1) US 11093557 B2(2021.08.17)
- [9] (특허문헌 2) US 11010371 B2(2021.05.18)

발명의 상세한 설명

기술적 과제

- [10] 본 발명은 개인별 혹은 팀별로 경험되는 소규모의 데이터에 기반한 개인화된 문자열-태그 매칭 구조의 개인 사전을 구축할 수 있는 서비스를 제공하는 시스템 및 방법을 제공하는 것을 목적으로 한다.

과제 해결 수단

- [11] 위와 같은 과제를 해결하기 위한 본 발명에 따른 개인 사전 구축 시스템은, 사용자별로 개인화된 문자열-태그 매칭을 통해 개인 사전을 구축하는 서비스를 제공하는 시스템으로서, 사용자가 태그를 작성하고자 하는 문자열을 선택하는 문자열 선택부, 상기 문자열 선택부에서 선택된 문자열을 서비스 공급자 서버로 전송

하는 문자열 송신부, 상기 서비스 공급자 서버로부터 전송받은 복수의 태그에서 사용자가 하나의 태그를 선택하거나 수동으로 입력할 수 있도록 하는 태그 선택 입력부를 포함하는 사용자 단말기; 및 상기 문자열 송신부로부터 문자열을 전송 받는 문자열 수신부, 상기 문자열 수신부에 수신된 문자열의 문자수를 카운트하여 제1 기준값과 비교 판단하는 제1 문자수 판단부, 상기 제1 문자수 판단부에서 문자수가 제1 기준값 이상인 것으로 판단되면 문자열을 요약한 요약문을 생성하는 요약문 추출부, 상기 문자열 수신부에 수신된 문자열의 문자수 또는 상기 요약문 추출부에서 생성된 요약문의 문자수를 카운트하여 제2 기준값과 비교 판단하는 제2 문자수 판단부, 상기 제2 문자수 판단부에서 문자수가 제2 기준값 미만인 것으로 판단되면 문자열 또는 요약문에 포함된 모든 단어를 추출하는 단어 추출부, 상기 제2 문자수 판단부에서 문자수가 제2 기준값 이상인 것으로 판단되면 문자열 또는 요약문에 포함된 주제를 추출하는 키워드 추출부, 상기 단어 추출부에서 추출된 단어 또는 상기 키워드 추출부에서 추출된 주제어에 상응하는 태그를 사용자 데이터베이스 또는 공급자 데이터베이스에서 추출하여 상기 사용자 단말기의 태그 선택입력부로 전송하는 태그 추천부, 상기 태그 선택입력부로부터 전송되는 태그를 상기 문자열과 매칭하여 사용자 데이터베이스에 저장하는 사전 저장부, 상기 사전 저장부에서 저장되는 태그의 스코어를 연산하여 사용자 데이터베이스에 저장하는 스코어 연산부를 포함하는 서비스 공급자 서버;를 포함하며, 상기 태그 추천부는 상기 사용자 데이터베이스에 저장된 스코어를 기준으로 높은 스코어를 갖는 상위 태그들을 추출하여 상기 사용자 단말기의 태그 선택입력부로 전송하고, 상기 스코어 연산부에서 태그의 스코어는 태그의 누적 선택 횟수, 수동 입력 여부, 최근 선택 이력, 사전 생성자 여부 및 공급자 데이터베이스 포함 여부를 변인으로 하여 연산되는 것을 특징으로 한다.

- [12] 바람직하게, 상기 스코어 연산부에서 스코어 연산시, 상기 스코어 변인들 중 태그의 누적 선택 횟수는 기본값이 되고, 나머지 스코어 변인들은 우선순위에 따라 상기 기본값에 부여되는 가중치로 작용할 수 있다.
- [13] 바람직하게, 상기 우선순위는 수동 입력 여부, 최근 선택 이력, 사전 생성자 여부, 공급자 데이터베이스 포함 여부 순으로 부여될 수 있다.
- [14] 또한, 상기 사용자 데이터베이스는, 단어 또는 주제어, 상기 단어 또는 주제어와 매칭된 태그 및 상기 태그의 스코어가 저장된 태그 추천 데이터베이스, 및 태그 및 상기 태그와 매칭된 문자열이 저장된 사전 데이터베이스를 포함한다.
- [15] 한편, 본 발명에 따른 개인 사전 구축 방법은, 사용자별로 개인화된 문자열-태그 매칭을 통해 개인 사전을 구축하는 서비스를 제공하는 방법으로서, 사용자 단말기 상에서 사용자가 태그를 작성하고자 하는 문자열을 선택하는 문자열 선택 단계; 상기 문자열 선택단계에서 선택된 문자열을 사용자 단말기에서 서비스 공급자 서버로 전송하는 문자열 송신단계; 상기 문자열 송신단계에서 전송된 문자열을 상기 서비스 공급자 서버로 수신하는 문자열 수신단계; 서비스 공급자 서버 상에서 상기 문자열 수신단계에서 수신된 문자열의 문자수를 카운트하여 제1 기

준값과 비교 판단하는 제1 문자수 판단단계; 서비스 공급자 서버 상에서 상기 제1 문자수 판단단계에서 문자수가 제1 기준값 이상인 것으로 판단되면 문자열을 요약한 요약문을 생성하는 요약문 추출단계; 서비스 공급자 서버 상에서 상기 문자열 수신단계에서 수신된 문자열의 문자수 또는 상기 요약문 추출단계에서 생성된 요약문의 문자수를 카운트하여 제2 기준값과 비교 판단하는 제2 문자수 판단단계; 서비스 공급자 서버 상에서 상기 제2 문자수 판단단계에서 문자수가 제2 기준값 미만인 것으로 판단되면 문자열 또는 요약문에 포함된 모든 단어를 추출하는 단어 추출단계; 서비스 공급자 서버 상에서 상기 제2 문자수 판단단계에서 문자수가 제2 기준값 이상인 것으로 판단되면 문자열 또는 요약문에 포함된 주제를 추출하는 키워드 추출단계; 서비스 공급자 서버 상에서 상기 단어 추출단계에서 추출된 단어 또는 상기 키워드 추출단계에서 추출된 주제에 상응하는 태그를 사용자 데이터베이스 또는 공급자 데이터베이스에서 추출하여 상기 사용자 단말기로 전송하는 태그 추천단계; 사용자 단말기 상에서 상기 서비스 공급자 서버로부터 전송받은 복수의 태그 중 사용자가 하나의 태그를 선택하거나 수동으로 입력하는 태그 선택입력단계; 상기 태그 선택입력단계에서 선택 또는 수동 입력된 태그를 서비스 공급자 서버 상에서 상기 문자열과 매칭하여 사용자 데이터베이스에 저장하는 사전 저장단계; 및 서비스 공급자 서버 상에서 상기 저장되는 태그의 스코어를 연산하여 사용자 데이터베이스에 저장하는 스코어 연산단계;를 포함하며, 상기 태그 추천단계는 상기 사용자 데이터베이스에 저장된 스코어를 기준으로 높은 스코어를 갖는 상위 태그들을 추출하여 상기 사용자 단말기로 전송하고, 상기 스코어 연산단계에서 태그의 스코어는 태그의 누적 선택 횟수, 수동 입력 여부, 최근 선택 이력, 사전 생성자 여부 및 공급자 데이터베이스 포함 여부를 변인으로 하여 연산되는 것을 특징으로 한다.

- [16] 바람직하게, 상기 스코어 연산단계에서 스코어 연산시, 상기 스코어 변인들 중 태그의 누적 선택 횟수는 기본값이 되고, 나머지 스코어 변인들은 우선순위에 따라 상기 기본값에 부여되는 가중치로 작용할 수 있다.
- [17] 바람직하게, 상기 우선순위는 수동 입력 여부, 최근 선택 이력, 사전 생성자 여부, 공급자 데이터베이스 포함 여부 순으로 부여될 수 있다.

발명의 효과

- [18] 본 발명에 따른 개인 사전 구축 시스템 및 방법은, 개인별 혹은 팀별로 경험되는 소규모의 데이터에 기반한 개인화된 문자열-태그 매칭 구조의 개인 사전을 구축할 수 있도록 함으로써, 사용자별로 개인화된 문구와 문장을 제공받을 수 있어 사용자별 경험에 기반한 문서를 보다 편리하고 효율적으로 작성할 수 있도록 한다.
- [19] 또한, 개인 사전 구축을 위한 태그를 추천함에 있어 태그의 누적 선택 횟수, 수동 입력 여부, 최근 선택 이력, 사전 생성자 여부 및 공급자 데이터베이스 포함 여

부를 변인으로 하여 결정된 스코어 순으로 상위 태그를 추천함으로써 사용자별로 개인화된 특성이 잘 반영될 수 있도록 한다.

- [20] 또한, 개인 사전 구축에 따라 문서 작성 서비스 제공시 발생할 수 있는 보안 문제, 저작권 문제 등을 해결할 수 있다.

도면의 간단한 설명

- [21] 도 1 및 도 2는 본 발명의 실시예에 따른 개인 사전 구축 시스템의 구성을 도식화하여 나타내는 도면이다.
- [22] 도 3은 본 발명의 실시예에 따른 개인 사전 구축 시스템에서 사용자 데이터베이스의 자료 구조를 도식화하여 나타내는 도면이다.
- [23] 도 4는 본 발명의 실시예에 따른 개인 사전 구축 방법을 개략적으로 나타내는 플로우 차트이다.
- [24] 도 5는 본 발명의 실시예에 따른 개인 사전 구축 방법의 과정을 세부적으로 나타내는 플로우 차트이다.
- [25] *도면 중 주요 부호에 대한 설명*
- [26] 100: 사용자 단말기 110: 문자열 선택부
- [27] 120: 문자열 송신부 130: 태그 선택입력부
- [28] 200: 서비스 공급자 서버 210: 문자열 수신부
- [29] 221: 제1 문자수 판단부 222: 제2 문자수 판단부
- [30] 223: 요약문 추출부 231: 단어 추출부
- [31] 232: 키워드 추출부 240: 태그 추천부
- [32] 250: 사전 저장부 260: 스코어 연산부
- [33] 271: 공급자 데이터베이스 272: 사용자 데이터베이스
- [34] 273: 공급자 데이터베이스 생성부 274: 태그 추천 데이터베이스
- [35] 275: 사전 데이터베이스
- [36] S100: 문자열 선택단계 S200: 문자열 송신단계
- [37] S300: 문자열 수신단계 S400: 제1 문자수 판단단계
- [38] S410: 요약문 추출단계 S500: 제2 문자수 판단단계
- [39] S610: 단어 추출단계 S620: 키워드 추출단계
- [40] S700: 태그 추천단계 S800: 태그 선택입력단계
- [41] S900: 사전 저장단계 S1000: 스코어 연산단계

발명의 실시를 위한 최선의 형태

- [42] 이하, 본 발명의 실시예에 따른 개인 사전 구축 시스템 및 방법을 첨부 도면을 참조하여 상세히 설명한다. 다만, 본 발명의 요지를 불필요하게 흐릴 수 있는 공지 기능 및 구성에 대한 상세한 설명은 생략하기로 한다.
- [43] 먼저, 본 발명의 실시예에 따른 개인 사전 구축 시스템에 대하여 도 1 내지 도 3을 참조하여 상세히 설명한다.

- [44] 도 1 및 도 2는 본 발명의 실시예에 따른 개인 사전 구축 시스템의 구성을 도식화하여 나타내는 도면이고, 도 3은 본 발명의 실시예에 따른 개인 사전 구축 시스템에서 사용자 데이터베이스의 자료 구조를 도식화하여 나타내는 도면이다.
- [45] 도 1에 보여지는 바와 같이, 본 발명의 실시예에 따른 개인 사전 구축 시스템은 사용자별로 개인화된 문자열-태그 매칭을 통해 개인 사전을 구축하는 서비스를 제공하는 시스템으로서, 크게 사용자 단말기(100) 및 서비스 공급자 서버(200)를 포함하여 이루어지며, 개인 사전 구축 서비스는 온라인 클라우드 플랫폼 상에서 인터넷 브라우저를 통해 SaaS(Software-as-a-Service) 기반의 소프트웨어로써 구현될 수 있다.
- [46] 도 2를 참조하면, 사용자 단말기(100)는 사용자가 태그를 작성하고자 하는 문자열을 선택하는 문자열 선택부(110), 상기 문자열 선택부(110)에서 선택된 문자열을 서비스 공급자 서버(200)로 전송하는 문자열 송신부(120), 상기 서비스 공급자 서버(200)로부터 전송받은 복수의 태그에서 사용자가 하나의 태그를 선택하거나 수동으로 입력할 수 있도록 하는 태그 선택입력부(130)를 포함한다.
- [47] 문자열 선택부(110)은 사용자 단말기(100)에 설치된 인터넷 브라우저의 웹확장 프로그램으로 구현되어 사용자가 인터넷 브라우저를 통해 접하는 신문 기사, 문서, 자료 등에서 특정 문구를 드래그하여 서비스 공급자 서버(200)로 전송할 문자열을 선택할 수 있다.
- [48] 또한, 문자열 선택부(110)은 사용자 단말기(100)에 설치된 인터넷 브라우저를 통해 접속되는 웹 기반의 클라우드 애플리케이션으로 구현되는 편집 프로그램을 통해 사용자가 특정 문구를 입력하고 드래그하여 서비스 공급자 서버(200)로 전송할 문자열을 선택할 수 있다.
- [49] 또한, 문자열 선택부(110)은 사용자 단말기(100)에 설치된 음성인식 프로그램으로 구현되어 사용자에게 의해 입력된 음성을 문자열로 변환하고 서비스 공급자 서버(200)로 전송할 문자열을 선택할 수 있다.
- [50] 문자열 송신부(120)는 사용자 단말기(100)에 내장된 통신 모듈로 구현되며, 유무선 통신을 통해 상기 문자열 선택부(110)에서 선택된 문자열을 서비스 공급자 서버(200)로 전송한다.
- [51] 태그 선택입력부(130)는 상기 서비스 공급자 서버(200)로 문자열 송신 후 상기 서비스 공급자 서버(200)로부터 전송받은 복수의 태그에서 사용자가 하나의 태그를 선택하거나 수동으로 입력할 수 있도록 한다.
- [52] 태그 선택입력부(130)에서 선택 또는 입력된 태그는 사용자 단말기(100)에 내장된 통신 모듈을 통해 서비스 공급자 서버(200)로 전송된다.
- [53] 태그 선택입력부(130)도 상기 문자열 선택부(110)와 마찬가지로 사용자 단말기(100)에 설치된 인터넷 브라우저의 웹확장 프로그램이나 클라우드 애플리케이션으로 구현되는 편집 프로그램 또는 음성인식 프로그램으로 구현될 수 있다.
- [54] 한편, 서비스 공급자 서버(200)는 상기 문자열 송신부(120)로부터 문자열을 전송받는 문자열 수신부(210), 상기 문자열 수신부(210)에 수신된 문자열의 문자수

를 카운트하여 제1 기준값과 비교 판단하는 제1 문자수 판단부(221), 상기 제1 문자수 판단부(221)에서 문자수가 제1 기준값 이상인 것으로 판단되면 문자열을 요약한 요약문을 생성하는 요약문 추출부(223), 상기 문자열 수신부(210)에 수신된 문자열의 문자수 또는 상기 요약문 추출부(223)에서 생성된 요약문의 문자수를 카운트하여 제2 기준값과 비교 판단하는 제2 문자수 판단부(222), 상기 제2 문자수 판단부(222)에서 문자수가 제2 기준값 미만인 것으로 판단되면 문자열 또는 요약문에 포함된 모든 단어를 추출하는 단어 추출부(231), 상기 제2 문자수 판단부(222)에서 문자수가 제2 기준값 이상인 것으로 판단되면 문자열 또는 요약문에 포함된 주제어를 추출하는 키워드 추출부(232), 상기 단어 추출부(231)에서 추출된 단어 또는 상기 키워드 추출부(232)에서 추출된 주제어에 상응하는 태그를 사용자 데이터베이스(272) 또는 공급자 데이터베이스(271)에서 추출하여 상기 사용자 단말기(100)의 태그 선택입력부(130)로 전송하는 태그 추천부(240), 상기 태그 선택입력부(130)로부터 전송되는 태그를 상기 문자열과 매칭하여 사용자 데이터베이스(272)에 저장하는 사전 저장부(250), 상기 사전 저장부(250)에서 저장되는 태그의 스코어를 연산하여 사용자 데이터베이스(272)에 저장하는 스코어 연산부(260)를 포함한다.

- [55] 문자열 수신부(210)는 서비스 공급자 서버(200)에 구비된 통신 모듈로 구현되며, 유무선 통신을 통해 상기 사용자 단말기(100)의 문자열 송신부(120)로부터 문자열을 전송받는다.
- [56] 제1 문자수 판단부(221), 요약문 추출부(223), 제2 문자수 판단부(222), 단어 추출부(231) 및 키워드 추출부(232)는 문자열-태그 매칭을 위해 문자열로부터 단어 또는 키워드를 추출하는 구성이다.
- [57] 제1 문자수 판단부(221)는 상기 문자열 수신부(210)에 수신된 문자열의 문자수를 카운트하여 제1 기준값과 비교 판단한다.
- [58] 제1 문자수 판단부(221)에서 카운트된 문자열의 문자수가 제1 기준값(예: 100개) 이상인 경우 요약문 추출부(223)는 요약문을 생성한다.
- [59] 요약문 추출부(223)는 주어진 문자열에서 중요도가 높거나 핵심이 되는 문장을 그대로 추출하는 추출적 요약(extractive summarization) 방법을 사용하는 것이 바람직하며, 텍스트랭크(TextRank)와 같은 공지의 다양한 요약문 추출 알고리즘이 사용될 수 있다.
- [60] 제2 문자수 판단부(222)는 상기 문자열 수신부(210)에 수신된 문자열의 문자수 또는 상기 요약문 추출부(223)에서 생성된 요약문의 문자수를 카운트하여 제2 기준값과 비교 판단한다.
- [61] 제2 문자수 판단부(222)에서 카운트된 문자수가 제2 기준값(예: 20개) 미만인 경우 단어 추출부(231)를 통해 문자열 또는 요약문에 포함된 모든 단어를 추출한다.

- [62] 반면, 제2 문자수 판단부(222)에서 카운트된 문자수가 제2 기준값(예: 20개) 이상인 경우에는 키워드 추출부(232)를 통해 문자열 또는 요약문에 포함된 주제어를 추출한다.
- [63] 단어 추출부(231)는 토큰화(tokenization), 정제(cleaning), 정규화(normalization) 등의 전처리 알고리즘을 사용하여 조사, 불용어 등의 불필요한 글자를 제거함으로써 문자열 또는 요약문에 포함된 모든 단어를 추출한다.
- [64] 키워드 추출부(232)는 카운트 기반 벡터화(CountVectorizer), 워드 임베딩(Word embedding) 등의 주제어 추출 알고리즘을 사용하여 주어진 문자열에서 중요도가 높거나 핵심이 되는 단어를 추출하며, 공지문의 다양한 키워드 추출 알고리즘이 사용될 수 있다.
- [65] 즉, 본 발명에서는 문자열-태그 매칭을 위해 문자열로부터 단어 또는 주제어를 추출함에 있어, 문자열의 문자수에 따른 요약문 추출, 주제어 추출, 단어 추출의 단계적인 프로세스를 구현함으로써 단어 또는 주제어 추출의 정확도 및 효율성을 제고할 수 있다.
- [66] 태그 추천부(240)는 상기 단어 추출부(231)에서 추출된 단어 또는 상기 키워드 추출부(232)에서 추출된 주제어에 상응하는 태그를 사용자 데이터베이스(272) 또는 공급자 데이터베이스(271)에서 추출하여 서비스 공급자 서버(200)에 구비된 통신 모듈을 통해 상기 사용자 단말기(100)의 태그 선택입력부(130)로 전송한다.
- [67] 이를 위해, 도 3을 참조하면, 사용자 데이터베이스(272)는 단어 또는 주제어, 상기 단어 또는 주제어와 매칭된 태그 및 상기 태그의 스코어가 저장된 태그 추천 데이터베이스(274)를 포함할 수 있다.
- [68] 예를 들어, 태그 추천 데이터베이스(274)에는 "JAVA"라는 단어 또는 주제어에 대해 "computer", "software", "HTML" 등의 태그가 매칭되어 저장되어 있고, 각 태그는 100점, 200점, 50점 등의 스코어가 저장되어 있다.
- [69] 태그 추천부(240)는 태그 추천 데이터베이스(274)에 저장된 스코어를 기준으로 높은 스코어를 갖는 상위 태그들(예: 상위 4개)을 추출하여 상기 사용자 단말기(100)의 태그 선택입력부(130)로 전송한다.
- [70] 사용자 데이터베이스(272)에 상기 단어 추출부(231)에서 추출된 단어 또는 상기 키워드 추출부(232)에서 추출된 주제어에 상응하는 태그가 없는 경우, 공급자 데이터베이스(271)로부터 태그를 추출한다.
- [71] 공급자 데이터베이스(271)는 모든 사용자가 이용할 수 있도록 공급자가 제공하는 범용 데이터베이스로서, 공급자 데이터베이스 생성부(273)에 의해 주기적으로 인터넷 자료를 무작위적으로 크롤링(crawling)하여 범용 태그를 생성하고 업데이트된 태그가 저장된다.
- [72] 사전 저장부(250)는 상기 태그 선택입력부(130)로부터 전송되는 태그를 상기 문자열과 매칭하여 사용자 데이터베이스(272)에 저장한다.

- [73] 이를 위해, 도 3를 참조하면, 사용자 데이터베이스(272)는 태그 및 태그와 매칭된 문자열이 저장되는 사전 데이터베이스(275)를 포함할 수 있다.
- [74] 이와 같이 태그 및 태그와 매칭된 문자열이 저장된 사전 데이터베이스(275)는 바로 본 발명에 따른 개인 사전 구축 시스템을 통해 사용자별로 개인화되어 구축되는 개인 사전이 된다.
- [75] 즉, 사용자는 사전 데이터베이스(275)를 통해 태그를 검색하면 태그와 매칭된 문자열을 제공 받을 수 있게 되고, 제공된 문자열을 문서 작성시 활용할 수 있게 된다.
- [76] 스코어 연산부(260)는 상기 태그 추천 데이터베이스(274)에 저장되는 태그의 스코어를 연산하여 업데이트한다.
- [77] 이때, 태그의 스코어는 태그의 누적 선택 횟수, 수동 입력 여부, 최근 선택 이력, 사전 생성자 여부 및 공급자 데이터베이스 포함 여부를 변인으로 하여 결정된다.
- [78] 태그의 누적 선택 횟수는 태그가 사용자에게 의해 선택된 누적 횟수이다. 누적 선택 횟수가 많을 수록 높은 점수를 부여하여 추천되는 태그들 중 상위에 위치될 수 있도록 한다.
- [79] 태그의 수동 입력 여부는 사용자가 추천된 태그를 선택하지 않고 수동으로 입력한 경우 개인 사전 구축에 있어 가장 높은 우선순위를 가지므로 가장 높은 점수를 부여하고, 사용자가 추천된 태그 중에 선택한 경우에는 다음 우선순위를 가지도록 점수를 부여한다. 선택적으로, 사용자가 추천된 태그 중에 부적합이라고 선택한 경우에는 감점하도록 할 수 있다.
- [80] 태그의 최근 선택 이력은 사용자에게 의해 최근에 선택된 태그일 수록 높은 점수를 부여하여 추천되는 태그들 중 상위에 위치될 수 있도록 한다.
- [81] 사전 생성자 여부는 개인 사전 구축에 여러 사용자가 참여할 경우, 사전을 처음 생성한 사용자가 선택한 태그에 높은 점수를 부여하여 추천되는 태그들 중 상위에 위치될 수 있도록 한다.
- [82] 이는 팀단위로 협업을 하는 경우 사전을 처음 생성한 사용자, 즉 팀리더가 작성한 태그가 팀원들이 작성한 태그에 비해 우선순위를 갖도록 하기 위함이다.
- [83] 공급자 데이터베이스 포함 여부는 태그가 공급자 데이터베이스(271)에 포함되어 있는 경우 최근 트렌드가 반영된 범용 태그임을 의미하므로 높은 점수를 부여하여 추천되는 태그들 중 상위에 위치될 수 있도록 한다.
- [84] 스코어 연산식 하기 식과 같이 상기 스코어 변인들 중 태그의 누적 선택 횟수는 기본값이 되고, 나머지 스코어 변인들은 우선순위에 따라 상기 기본값에 부여되는 가중치로 작용할 수 있다.
- [85]
- [86] (스코어 연산식)
- [87]
$$\text{스코어} = \text{기존스코어} + \text{기본값} \times \text{가중치}_1 \times \text{가중치}_2 \times \dots \times \text{가중치}_n - \text{감점}$$
- [88]

- [89] 상기 식에서 감점은 사용자가 추천된 태그 중에 부적합이라고 선택한 경우에 부여되는 점수이다.
- [90] 본 발명의 실시예에 따른 개인 사전 구축 시스템에서는 수동 입력 여부, 최근 선택 이력, 사전 생성자 여부, 공급자 데이터베이스 포함 여부 순으로 가중치를 부여한다.
- [91] 예를 들어, 태그의 누적 선택 횟수에 따른 기본값은 하나의 단어 또는 주제어에 매칭된 모든 태그들의 누적 선택 횟수를 시그모이드(sigmoid) 함수를 이용하여 0부터 1사이의 값으로 정규화(normalize)하여 부여될 수 있다.
- [92] 가중치1은 수동 입력 여부에 따라 수동 입력된 태그인 경우 2, 추천된 태그 중에 선택된 태그인 경우 1.5, 추천된 태그 중에 선택되지 않은 태그인 경우 1.0의 값이 부여될 수 있다.
- [93] 가중치2는 최근 선택 이력에 따라 최근 선택 순으로 1.4부터 1.0사이의 값이 부여될 수 있다.
- [94] 가중치3은 사전 생성자 여부에 따라 사용자가 사전을 처음 생성했던 사전 생성자인 경우 1.3, 그렇지 않은 경우 1.0의 값이 부여될 수 있다.
- [95] 가중치4는 공급자 데이터베이스 포함 여부에 따라 태그가 공급자 데이터베이스(271)에 포함되어 있는 경우 1.1, 그렇지 않은 경우 1.0의 값이 부여될 수 있다.
- [96] 감점은 사용자가 추천된 태그 중에 부적합이라고 선택한 경우, 0.1의 값이 부여될 수 있다.
- [97] 또한, 상기 식에서 기존스코어도 하나의 단어 또는 주제어에 매칭된 모든 태그들의 기존스코어를 시그모이드(sigmoid) 함수를 이용하여 0부터 1사이의 값으로 정규화(normalize)하여 부여될 수 있다. 이는, 오래된 태그가 누적된 스코어로 인해 새로운 최신 태그에 비해 높은 스코어를 계속 유지하게 되는 오류를 방지하기 위함이다.
- [98] 위와 같이 본 발명의 실시예에 따른 개인 사전 구축 시스템은 개인 사전 구축을 위한 태그를 추천함에 있어 태그의 누적 선택 횟수, 수동 입력 여부, 최근 선택 이력, 사전 생성자 여부 및 공급자 데이터베이스 포함 여부를 변인으로 하여 결정된 스코어 순으로 상위 태그를 추천함으로써 사용자별로 개인화된 특성이 잘 반영될 수 있도록 한다.
- [99] 다음으로, 도 4 및 도 5를 참조하여 본 발명의 실시예에 따른 개인 사전 구축 방법에 대하여 상세히 설명한다.
- [100] 도 4는 본 발명의 실시예에 따른 개인 사전 구축 방법을 개략적으로 나타내는 플로우 차트이고, 도 5는 본 발명의 실시예에 따른 개인 사전 구축 방법의 과정을 세부적으로 나타내는 플로우 차트이다.
- [101] 본 발명의 실시예에 따른 개인 사전 구축 방법은, 도 4 및 도 5에 나타낸 바와 같이, 문자열 선택단계(S100), 문자열 송신단계(S200), 문자열 수신단계(S300), 제1 문자수 판단단계(S400), 요약문 추출단계(S410), 제2 문자수 판단단계(S500), 단

어 추출단계(S610), 키워드 추출단계(S620), 태그 추천단계(S700), 태그 선택입력 단계(S800), 사전 저장단계(S900), 및 스코어 연산단계(S1000)를 포함한다.

- [102] 구체적으로, 본 발명의 실시예에 따른 개인 사전 구축 방법은, 사용자별로 개인화된 문자열-태그 매칭을 통해 개인 사전을 구축하는 서비스를 제공하는 방법으로서, 사용자 단말기(100) 상에서 사용자가 태그를 작성하고자 하는 문자열을 선택하는 문자열 선택단계(S100), 상기 문자열 선택단계(S100)에서 선택된 문자열을 사용자 단말기(100)에서 서비스 공급자 서버(200)로 전송하는 문자열 송신단계(S200), 상기 문자열 송신단계(S200)에서 전송된 문자열을 상기 서비스 공급자 서버(200)로 수신하는 문자열 수신단계(S300), 서비스 공급자 서버(200) 상에서 상기 문자열 수신단계(S300)에서 수신된 문자열의 문자수를 카운트하여 제1 기준값과 비교 판단하는 제1 문자수 판단단계(S400), 서비스 공급자 서버(200) 상에서 상기 제1 문자수 판단단계(S400)에서 문자수가 제1 기준값 이상인 것으로 판단되면 문자열을 요약한 요약문을 생성하는 요약문 추출단계(S410), 서비스 공급자 서버(200) 상에서 상기 문자열 수신단계(S300)에서 수신된 문자열의 문자수 또는 상기 요약문 추출단계(S410)에서 생성된 요약문의 문자수를 카운트하여 제2 기준값과 비교 판단하는 제2 문자수 판단단계(S500), 서비스 공급자 서버(200) 상에서 상기 제2 문자수 판단단계(S500)에서 문자수가 제2 기준값 미만인 것으로 판단되면 문자열 또는 요약문에 포함된 모든 단어를 추출하는 단어 추출단계(S610), 서비스 공급자 서버(200) 상에서 상기 제2 문자수 판단단계(S500)에서 문자수가 제2 기준값 이상인 것으로 판단되면 문자열 또는 요약문에 포함된 주제를 추출하는 키워드 추출단계(S620), 서비스 공급자 서버(200) 상에서 상기 단어 추출단계(S610)에서 추출된 단어 또는 상기 키워드 추출단계(S620)에서 추출된 주제어에 상응하는 태그를 사용자 데이터베이스(272) 또는 공급자 데이터베이스(271)에서 추출하여 상기 사용자 단말기(100)로 전송하는 태그 추천단계(S700), 사용자 단말기(100) 상에서 상기 서비스 공급자 서버(200)로부터 전송받은 복수의 태그 중 사용자가 하나의 태그를 선택하거나 수동으로 입력하는 태그 선택입력단계(S800), 상기 태그 선택입력단계(S800)에서 선택 또는 수동 입력된 태그를 서비스 공급자 서버(200) 상에서 상기 문자열과 매칭하여 사용자 데이터베이스(272)에 저장하는 사전 저장단계(S900), 및 서비스 공급자 서버(200) 상에서 상기 저장되는 태그의 스코어를 연산하여 사용자 데이터베이스(272)에 저장하는 스코어 연산단계(S1000)를 포함한다.

- [103] 상기 문자열 선택단계(S100)는 사용자 단말기(100) 상에서 사용자가 태그를 작성하고자 하는 문자열을 선택하는 단계이다.

- [104] 여기에서, 사용자는 사용자 단말기(100) 상에서 인터넷 브라우저를 통해 SaaS(Software-as-a-Service) 기반의 소프트웨어로 구현되는 개인 사전 구축 서비스에 접속하여 서비스 공급자 서버(200)에 로그인하고, 사용자 단말기(100)에 설치된 인터넷 브라우저의 웹확장 프로그램이나 클라우드 애플리케이션으로 구현

되는 편집 프로그램 또는 음성인식 프로그램으로 이루어진 문자열 선택부(110)를 통해 문자열을 선택할 수 있다.

- [105] 로그인시, 사용자가 구축한 개인 사전을 불러올 수 있으며, 사용자가 타인이 구축 중인 개인 사전에 참여할 경우 참여하고 있는 타인의 개인 사전도 불러올 수 있다.
- [106] 계속해서, 문자열 송신단계(S200)는 상기 문자열 선택단계(S100)에서 선택된 문자열을 문자열 송신부(120)를 통해 유무선 통신으로 사용자 단말기(100)에서 서비스 공급자 서버(200)로 전송하는 단계이다.
- [107] 그리고, 문자열 수신단계(S300)는 상기 문자열 송신단계(S200)에서 전송된 문자열을 문자열 수신부(210)를 통해 유무선 통신으로 상기 서비스 공급자 서버(200)로 수신하는 단계이다.
- [108] 계속해서, 제1 문자수 판단단계(S400)는 서비스 공급자 서버(200) 상에서 제1 문자수 판단부(221)에 의해 상기 문자열 수신단계(S300)에서 수신된 문자열의 문자수를 카운트하여 제1 기준값과 비교 판단하는 단계이다.
- [109] 이때, 제1 문자수 판단단계(S400)에서 카운트된 문자열의 문자수가 제1 기준값(예: 100개) 이상인 경우, 요약문 추출단계(S410)를 수행하여 문자열을 요약한 요약문을 생성한다. 요약문 추출단계(S410)는 요약문 추출부(223)를 통해 수행된다.
- [110] 그리고, 제2 문자수 판단단계(S500)는 서비스 공급자 서버(200) 상에서 제2 문자수 판단부(222)에 의해 상기 문자열 수신단계(S300)에서 수신된 문자열의 문자수 또는 상기 요약문 추출단계(S410)에서 생성된 요약문의 문자수를 카운트하여 제2 기준값과 비교 판단하는 단계이다.
- [111] 이때, 상기 제2 문자수 판단단계(S500)에서 문자수가 제2 기준값 미만인 것으로 판단되면 단어 추출단계(S610)가 수행되고, 상기 제2 문자수 판단단계(S500)에서 문자수가 제2 기준값 이상인 것으로 판단되면 키워드 추출단계(S620)가 수행된다.
- [112] 여기서, 단어 추출단계(S610)는 서비스 공급자 서버(200) 상에서 단어 추출부(231)에 의해 문자열 또는 요약문에 포함된 모든 단어를 추출하는 단계이다.
- [113] 또한, 키워드 추출단계(S620)는 서비스 공급자 서버(200) 상에서 키워드 추출부(232)에 의해 문자열 또는 요약문에 포함된 주제어를 추출하는 단계이다.
- [114] 계속해서, 태그 추천단계(S700)는 서비스 공급자 서버(200) 상에서 태그 추천부(240)에 의해 상기 단어 추출단계(S610)에서 추출된 단어 또는 상기 키워드 추출단계(S620)에서 추출된 주제어에 상응하는 태그를 사용자 데이터베이스(272) 또는 공급자 데이터베이스(271)에서 추출하여 상기 사용자 단말기(100)로 전송하는 단계이다.
- [115] 이때, 태그 추천부(240)는 사용자 데이터베이스(272)의 태그 추천 데이터베이스(274)에 저장된 스코어를 기준으로 높은 스코어를 갖는 상위 태그들(예: 상위 4개)을 추출한다.

- [116] 태그의 스코어는 태그의 누적 선택 횟수, 수동 입력 여부, 최근 선택 이력, 사전 생성자 여부 및 공급자 데이터베이스 포함 여부를 변인으로 하여 결정된다.
- [117] 계속해서, 태그 선택입력단계(S800)는 사용자 단말기(100) 상에서 태그 선택입력부(130)를 통해 상기 서비스 공급자 서버(200)로부터 전송받은 복수의 태그 중 사용자가 하나의 태그를 선택하거나 수동으로 입력하는 단계이다.
- [118] 이때, 사용자 단말기(100) 상에는 사용자가 수동으로 태그를 입력할 수 있는 입력창과 함께 추천된 태그들에 대해 선택/부적합 정보를 입력할 수 있는 선택창이 띄워진다.
- [119] 그리고, 사전 저장단계(S900)는 상기 태그 선택입력단계(S800)에서 선택 또는 수동 입력된 태그를 서비스 공급자 서버(200) 상에서 사전 저장부(250)에 의해 상기 문자열과 매칭하여 사용자 데이터베이스(272)에 저장하는 단계이다.
- [120] 계속해서, 스코어 연산단계(S1000)는 서비스 공급자 서버(200) 상에서 스코어 연산부(260)에 의해 상기 저장되는 태그의 스코어를 연산하여 사용자 데이터베이스(272)에 저장하는 단계이다.
- [121] 스코어는 상기 스코어 변인들 중 태그의 누적 선택 횟수는 기본값이 되고, 나머지 스코어 변인들은 우선순위에 따라 상기 기본값에 부여되는 가중치로 하여 연산될 수 있다. 선택적으로, 사용자가 추천된 태그 중에 부적합이라고 선택한 경우 스코어 연산시 감점으로 적용할 수 있다.
- [122] 위와 같이 본 발명의 실시예에 따른 개인 사전 구축 방법은 개인 사전 구축을 위한 태그를 추천함에 있어 태그의 누적 선택 횟수, 수동 입력 여부, 최근 선택 이력, 사전 생성자 여부 및 공급자 데이터베이스 포함 여부를 변인으로 하여 결정된 스코어 순으로 상위 태그를 추천함으로써 사용자별로 개인화된 특성이 잘 반영될 수 있도록 한다.
- [123] 이상, 본 발명의 실시예를 첨부된 도면을 참조하여 상세하게 설명하였으나, 본 명세서와 첨부된 도면에 개시된 실시예들은 본 발명의 기술적 사상을 쉽게 설명하기 위한 목적으로 사용된 것일 뿐, 특허청구범위에 기재된 본 발명의 범위를 제한하기 위하여 사용된 것은 아니며, 따라서 본 기술분야의 통상의 지식을 가진 자라면 이로부터 다양한 변형 및 균등한 타 실시예가 가능하다는 점을 이해할 것이다.

청구범위

- [청구항 1] 사용자별로 개인화된 문자열-태그 매칭을 통해 개인 사전을 구축하는 서비스를 제공하는 시스템으로서,
 사용자가 태그를 작성하고자 하는 문자열을 선택하는 문자열 선택부, 상기 문자열 선택부에서 선택된 문자열을 서비스 공급자 서버로 전송하는 문자열 송신부, 상기 서비스 공급자 서버로부터 전송받은 복수의 태그에서 사용자가 하나의 태그를 선택하거나 수동으로 입력할 수 있도록 하는 태그 선택입력부를 포함하는 사용자 단말기; 및
 상기 문자열 송신부로부터 문자열을 전송받는 문자열 수신부, 상기 문자열 수신부에 수신된 문자열의 문자수를 카운트하여 제1 기준값과 비교 판단하는 제1 문자수 판단부, 상기 제1 문자수 판단부에서 문자수가 제1 기준값 이상인 것으로 판단되면 문자열을 요약한 요약문을 생성하는 요약문 추출부, 상기 문자열 수신부에 수신된 문자열의 문자수 또는 상기 요약문 추출부에서 생성된 요약문의 문자수를 카운트하여 제2 기준값과 비교 판단하는 제2 문자수 판단부, 상기 제2 문자수 판단부에서 문자수가 제2 기준값 미만인 것으로 판단되면 문자열 또는 요약문에 포함된 모든 단어를 추출하는 단어 추출부, 상기 제2 문자수 판단부에서 문자수가 제2 기준값 이상인 것으로 판단되면 문자열 또는 요약문에 포함된 주제어를 추출하는 키워드 추출부, 상기 단어 추출부에서 추출된 단어 또는 상기 키워드 추출부에서 추출된 주제어에 상응하는 태그를 사용자 데이터베이스 또는 공급자 데이터베이스에서 추출하여 상기 사용자 단말기의 태그 선택입력부로 전송하는 태그 추천부, 상기 태그 선택입력부로부터 전송되는 태그를 상기 문자열과 매칭하여 사용자 데이터베이스에 저장하는 사전 저장부, 상기 사전 저장부에서 저장되는 태그의 스코어를 연산하여 사용자 데이터베이스에 저장하는 스코어 연산부를 포함하는 서비스 공급자 서버;
 를 포함하며,
 상기 태그 추천부는 상기 사용자 데이터베이스에 저장된 스코어를 기준으로 높은 스코어를 갖는 상위 태그들을 추출하여 상기 사용자 단말기의 태그 선택입력부로 전송하고,
 상기 스코어 연산부에서 태그의 스코어는 태그의 누적 선택 횟수, 수동 입력 여부, 최근 선택 이력, 사전 생성자 여부 및 공급자 데이터베이스 포함 여부를 변인으로 하여 연산되는 것을 특징으로 하는 개인 사전 구축 시스템.
- [청구항 2] 청구항 1에 있어서,
 상기 스코어 연산부에서 스코어 연산시, 상기 스코어 변인들 중 태그의 누적 선택 횟수는 기본값이 되고, 나머지 스코어 변인들은 우선순위에 따라

상기 기본값에 부여되는 가중치로 작용하는 것을 특징으로 하는 개인 사전 구축 시스템.

[청구항 3]

청구항 2에 있어서,

상기 우선순위는 수동 입력 여부, 최근 선택 이력, 사전 생성자 여부, 공급자 데이터베이스 포함 여부 순으로 부여되는 것을 특징으로 하는 개인 사전 구축 시스템.

[청구항 4]

청구항 1에 있어서,

상기 사용자 데이터베이스는,

단어 또는 주제어, 상기 단어 또는 주제어와 매칭된 태그 및 상기 태그의 스코어가 저장된 태그 추천 데이터베이스, 및 태그 및 상기 태그와 매칭된 문자열이 저장된 사전 데이터베이스를 포함하는 것을 특징으로 하는 개인 사전 구축 시스템.

[청구항 5]

사용자별로 개인화된 문자열-태그 매칭을 통해 개인 사전을 구축하는 서비스를 제공하는 방법으로서,

사용자 단말기 상에서 사용자가 태그를 작성하고자 하는 문자열을 선택하는 문자열 선택단계;

상기 문자열 선택단계에서 선택된 문자열을 사용자 단말기에서 서비스 공급자 서버로 전송하는 문자열 송신단계;

상기 문자열 송신단계에서 전송된 문자열을 상기 서비스 공급자 서버로 수신하는 문자열 수신단계;

서비스 공급자 서버 상에서 상기 문자열 수신단계에서 수신된 문자열의 문자수를 카운트하여 제1 기준값과 비교 판단하는 제1 문자수 판단단계;

서비스 공급자 서버 상에서 상기 제1 문자수 판단단계에서 문자수가 제1 기준값 이상인 것으로 판단되면 문자열을 요약한 요약문을 생성하는 요약문 추출단계;

서비스 공급자 서버 상에서 상기 문자열 수신단계에서 수신된 문자열의 문자수 또는 상기 요약문 추출단계에서 생성된 요약문의 문자수를 카운트하여 제2 기준값과 비교 판단하는 제2 문자수 판단단계;

서비스 공급자 서버 상에서 상기 제2 문자수 판단단계에서 문자수가 제2 기준값 미만인 것으로 판단되면 문자열 또는 요약문에 포함된 모든 단어를 추출하는 단어 추출단계;

서비스 공급자 서버 상에서 상기 제2 문자수 판단단계에서 문자수가 제2 기준값 이상인 것으로 판단되면 문자열 또는 요약문에 포함된 주제어를 추출하는 키워드 추출단계;

서비스 공급자 서버 상에서 상기 단어 추출단계에서 추출된 단어 또는 상기 키워드 추출단계에서 추출된 주제어에 상응하는 태그를 사용자 데이터베이스 또는 공급자 데이터베이스에서 추출하여 상기 사용자 단말기로 전송하는 태그 추천단계;

사용자 단말기 상에서 상기 서비스 공급자 서버로부터 전송받은 복수의 태그 중 사용자가 하나의 태그를 선택하거나 수동으로 입력하는 태그 선택입력단계;

상기 태그 선택입력단계에서 선택 또는 수동 입력된 태그를 서비스 공급자 서버 상에서 상기 문자열과 매칭하여 사용자 데이터베이스에 저장하는 사전 저장단계; 및

서비스 공급자 서버 상에서 상기 저장되는 태그의 스코어를 연산하여 사용자 데이터베이스에 저장하는 스코어 연산단계;를 포함하며,

상기 태그 추천단계는 상기 사용자 데이터베이스에 저장된 스코어를 기준으로 높은 스코어를 갖는 상위 태그들을 추출하여 상기 사용자 단말기로 전송하고,

상기 스코어 연산단계에서 태그의 스코어는 태그의 누적 선택 횟수, 수동 입력 여부, 최근 선택 이력, 사전 생성자 여부 및 공급자 데이터베이스 포함 여부를 변인으로 하여 연산되는 것을 특징으로 하는 개인 사전 구축 방법.

[청구항 6]

청구항 5에 있어서,

상기 스코어 연산단계에서 스코어 연산시, 상기 스코어 변인들 중 태그의 누적 선택 횟수는 기본값이 되고, 나머지 스코어 변인들은 우선순위에 따라 상기 기본값에 부여되는 가중치로 작용하는 것을 특징으로 하는 개인 사전 구축 방법.

[청구항 7]

청구항 6에 있어서,

상기 우선순위는 수동 입력 여부, 최근 선택 이력, 사전 생성자 여부, 공급자 데이터베이스 포함 여부 순으로 부여되는 것을 특징으로 하는 개인 사전 구축 방법.