

(12) 특허협력조약에 의하여 공개된 국제출원

(19) 세계지식재산권기구
국제사무국

(43) 국제공개일

2025년 5월 1일 (01.05.2025)



(10) 국제공개번호

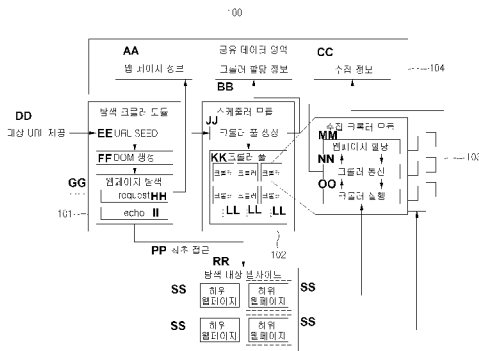
WO 2025/089616 A1

- (51) 국제특허분류:
G06F 16/951 (2019.01) G06F 9/48 (2006.01)
G06F 16/955 (2019.01) G06F 9/50 (2006.01)
- (21) 국제출원번호: PCT/KR2024/013922
- (22) 국제출원일: 2024년 9월 12일 (12.09.2024)
- (25) 출원언어: 한국어
- (26) 공개언어: 한국어
- (30) 우선권정보:
10-2023-0143615 2023년 10월 25일 (25.10.2023) KR
- (71) 출원인: 세종대학교산학협력단 (INDUSTRY-ACADEMIA COOPERATION GROUP OF SEJONG UNIVERSITY) [KR/KR]; 05006 서울특별시 광진구 능동로 209 (KR).
- (72) 발명자: 박기웅 (PARK, Ki-Woong); 05006 서울특별시 광진구 능동로 209 (KR). 안성규 (AHN, Sung-Kyu); 05006 서울특별시 광진구 능동로 209 (KR). 배승진 (BEA, Seung-Jin); 05006 서울특별시 광진구 능동로 209 (KR).

- (74) 대리인: 양정보 (YANG, Sungbo); 06099 서울특별시 강남구 선릉로125길 14 삼성빌딩 2층 피앤티특허법률사무소 (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, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, 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, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.
- (84) 지정국 (별도의 표시가 없는 한, 가능한 모든 종류의 역내 권리의 보호를 위하여): ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, 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, ME, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM,

(54) Title: CONFIDENTIAL AND ROBUST DATA COLLECTION METHOD AND SYSTEM VIA FLEXIBLE DISTRIBUTED CONTROL TECHNIQUE FOR DISTRIBUTED CRAWLING IN SWAM FORMAT

(54) 발명의 명칭: SWAM 형태의 분산 크롤링에 대한 유동적 분산 제어기법을 통한 기밀-강인 데이터 수집 방법 및 시스템



101 ... Search crawler module
102 ... Scheduler module
103 ... Collection crawler module
104 ... Shared data area
AA ... Web page information
BB ... Crawler assignment information
CC ... Collection information
DD ... Provide target URL
EE ... URL SEED
FF ... Generate DOM
GG ... Search web page
HH ... Request
II ... Echo
JJ ... Generate crawler pool
KK ... Crawler pool
LL ... Crawler
MM ... Assign web page
NN ... Crawler communication
OO ... Execute crawler
PP ... Initial access
RR ... Web site to be searched
SS ... Sub-web page

(57) Abstract: Disclosed are a confidential and robust data collection method and system via a flexible distributed control technique for distributed crawling in a SWAM format. A data collection method performed by a crawler according to an embodiment comprises the steps of: configuring multiple modularized crawler modules in order to perform a single function of crawling; and performing parallel crawling via web page access using the configured multiple crawler modules.

(57) 요약서: Swam 형태의 분산 크롤링에 대한 유동적 분산 제어기법을 통한 기밀-강인 데이터 수집 방법 및 시스템이 개시된다. 일 실시예에 따른 크롤러에 의해 수행되는 데이터 수집 방법은, 크롤링의 단일 기능을 수행하기 위해 모듈화된 다중 크롤러 모듈을 구성하는 단계; 및 상기 구성된 다중 크롤러 모듈을 이용한 웹 페이지 접속을 통해 병렬적 크롤링을 수행하는 단계를 포함할 수 있다.

WO 2025/089616 A1

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: SWAM 형태의 분산 크롤링에 대한 유동적 분산 제어기법을 통한 기밀-강인 데이터 수집 방법 및 시스템 기술분야

[1] 아래의 설명은 데이터 크롤링 기술에 관한 것이다.

[2]

배경기술

[3]웹의 규모가 폭발적으로 증가하면서 웹 페이지의 양 또한 기하급수적으로 늘어나고 있다. 초기에는 단일 크롤러를 사용해 웹을 순회하며 정보를 수집하는 중앙 집중형 구조가 주를 이루었다. 그러나, 현재 인터넷의 규모와 복잡성을 고려하면, 단일 크롤러만으로는 처리해야 할 웹 페이지의 양과 다양성에 어려움을 겪게 된다. 이로 인해 대규모 웹 데이터 수집에는 많은 인력과 시간이 소요되며, 효율적인 크롤링이 어려워지고 있다.

[4]또한, 웹 크롤링 기술의 발전과 함께 웹 페이지를 보유한 기관이나 단체들은 자신들의 정보가 유출되는 것을 방지하려는 시도를 활발히 하고 있다. 이에 대한 대표적인 예로, 웹 크롤러에 의한 트래픽 과부하나 너무 잦은 접속을 차단하는 조치가 있다. 이러한 환경 속에서 특히 악성코드 정보와 같은 민감한 데이터를 크롤링하는 것은 더욱 어려운 일이다. 악성코드 정보는 그 자체로 중요한 정보이며, 이를 악의를 가진 사용자에게 의해 심각한 문제를 초래할 수 있기 때문에, 웹 페이지 소유자들은 해당 정보에 대한 크롤링을 특히 제한하곤 한다.

[5]따라서, 현대의 웹 크롤링 환경에서는 다중 분산 크롤러와 같은 새로운 접근법이 필요하며, 특히 악성코드 정보와 같은 민감한 데이터 크롤링에는 더욱 신중하고 전략적인 방법이 요구된다.

[6]일례로, 대한민국 공개특허 10-2020-0019288(2020.02.24. 공개일)에 JCA(Java Client Application), JSA(Java Server Application), DB 서버, 웹서버로 구성되며, JCA와 JSA에 각자의 독립된 스레드를 세 개 이상 나누어 작업을 실행함으로써, 독립된 환경에서의 클라이언트가 큐와 스레드를 이용하여 빠르고 신뢰성이 향상된 웹 문서를 제공하는 다중 스레드 방식의 웹 크롤링 시스템이 개시된 바 있다.

[7]

발명의 상세한 설명

기술적 과제

[8]크롤러의 기능을 모듈화함으로써, 기존 단일 크롤링에서 수행되는 크롤링을 수행하는 연산을 최소 단위로 분할하여 다중화하고, 각 모듈에게 별도의 크롤링 타겟을 부여함으로써, 각각의 모듈이 웹 페이지에 접속하여 상호 교류를 통해 능동

적으로 정보를 수집하고, 웹 페이지의 접근 차단으로 인한 크롤링 기능의 중단을 방지하기 위해 다중 노드로 구성된 서버 및 클라우드 환경을 이용한다.

- [9] 각각의 크롤링 모듈이 서로 다른 웹 페이지를 대상으로 크롤링 기능을 수행하여 웹 페이지의 차단을 회피하고 다양한 정보를 병렬적으로 수집함으로써, 빠른 크롤링이 가능하도록 하여 웹 페이지로부터의 크롤링 차단 기능을 우회하고 빠른 크롤링이 가능할 수 있도록 한다.

[10]

과제 해결 수단

- [11] 데이터 수집 시스템에 의해 수행되는 데이터 수집 방법은, 크롤링의 단일 기능을 수행하기 위해 모듈화된 다중 크롤러 모듈을 구성하는 단계; 및 상기 구성된 다중 크롤러 모듈을 이용한 웹 페이지 접속을 통해 병렬적 크롤링을 수행하는 단계를 포함할 수 있다.
- [12] 상기 다중 크롤러 모듈은, 웹 페이지 내 크롤링에 요구되는 웹 페이지 정보를 탐색하는 탐색 크롤러 모듈, 상기 탐색된 웹 페이지 정보에 기초하여 작업 공간을 동적으로 할당하는 스케줄러 모듈 및 상기 할당된 작업 공간에 기초하여 크롤링 수행하는 수집 크롤러 모듈을 포함할 수 있다.
- [13] 상기 병렬적 크롤링을 수행하는 단계는, 탐색 크롤러 모듈에서, 사용자에게 전달받은 URL 정보를 기반으로 DOM(Document Object Model) 구조를 생성하고, 리퀘스트(Request)와 에코(echo) 패킷을 전송함에 따라 콘텐츠의 크기(content_length)와 ping 명령어의 응답속도를 측정하는 단계를 포함할 수 있다.
- [14] 상기 병렬적 크롤링을 수행하는 단계는, 상기 측정된 응답속도에 기초하여 전체 비율 1에서 상기 ping 명령어의 응답속도에 콘텐츠의 크기를 나눈 값을 감산하여 웹 페이지 구성요소의 크기 및 크롤링을 위한 가중치를 산정하는 단계를 포함할 수 있다.
- [15] 상기 병렬적 크롤링을 수행하는 단계는, 상기 리퀘스트(Request)와 에코(echo) 패킷을 전송함에 따라 탐색된 웹 페이지 구성요소 및 하위 페이지를 포함하는 웹 페이지 정보를 공유 데이터 영역에 저장하는 단계를 포함할 수 있다.
- [16] 상기 병렬적 크롤링을 수행하는 단계는, 스케줄러 모듈에서, 탐색 크롤러 모듈로부터 전달된 가중치를 기반으로 크롤러 풀을 이용하여 수집 크롤러 모듈의 개수를 연산하고, 상기 연산된 수집 크롤러 모듈의 개수에 기초하여 수집 크롤러 모듈을 생성하는 단계를 포함할 수 있다.
- [17] 상기 병렬적 크롤링을 수행하는 단계는, 상기 연산된 수집 크롤러 모듈의 개수에 기초하여 생성된 수집 크롤러 모듈을 작업 공간에 할당하고, 상기 생성된 수집 크롤러 모듈에게 상기 할당된 작업 공간이 가리키는 작업 정보를 전달하는 단계를 포함할 수 있다.

- [18] 상기 병렬적 크롤링을 수행하는 단계는, 수집 크롤러 모듈에서, 스케줄러 모듈에 의해 할당된 작업 공간의 웹 페이지 구성요소 및 하위 페이지 정보를 포함하는 작업 정보를 통해 크롤링을 수행하는 단계를 포함할 수 있다.
- [19] 상기 병렬적 크롤링을 수행하는 단계는, 상기 작업 공간에 연결되어 있는 진행 상황을 이용하여 이어서 진행할 작업이나 수집 크롤러 모듈의 작업 공간을 확인한 후, 크롤링을 수행하는 단계를 포함할 수 있다.
- [20] 상기 병렬적 크롤링을 수행하는 단계는, 상기 수행된 크롤링이 완료되었는지 여부를 판단하고, 상기 수행된 크롤링이 완료되지 않았을 경우 크롤링 진행 상황을 업데이트하고, 상기 수행된 크롤링이 완료되었을 경우 스케줄러 모듈이 다음 작업 공간을 할당하였는지 여부를 판단하는 단계를 포함할 수 있다.
- [21] 상기 병렬적 크롤링을 수행하는 단계는, 상기 스케줄러 모듈이 다음 작업 공간을 할당하지 않은 것으로 판단됨에 따라 크롤링 진행 상황을 업데이트하고 크롤러 대기열로 이동하는 단계를 포함할 수 있다.
- [22] 상기 병렬적 크롤링을 수행하는 단계는, 상기 스케줄러 모듈이 다음 작업 공간을 할당한 것으로 판단됨에 따라 수집 크롤러 모듈을 할당된 다음 작업 공간에 배치하는 단계를 포함할 수 있다.
- [23] 상기 병렬적 크롤링을 수행하는 단계는, 상기 수행된 크롤링을 통해 수집된 정보를 공유 데이터 영역에 저장하는 단계를 포함할 수 있다.
- [24] 데이터 수집 방법을 상기 데이터 수집 시스템에 실행시키기 위해 비-일시적인 컴퓨터 판독가능한 기록 매체에 저장되는 컴퓨터 프로그램을 포함할 수 있다.
- [25] 데이터 수집 시스템에 있어서, 메모리에 포함된 컴퓨터 판독가능한 명령들을 실행하도록 구성된 적어도 하나의 프로세서를 포함하고, 상기 적어도 하나의 프로세서는, 크롤링의 단일 기능을 수행하기 위해 모듈화된 다중 크롤러 모듈을 구성하고, 상기 구성된 다중 크롤러 모듈을 이용한 웹 페이지 접속을 통해 병렬적 크롤링을 수행할 수 있다.

[26]

발명의 효과

- [27] 웹 페이지에서 정보를 수집하는 과정에서 지속적인 크롤링 기능의 수행으로 인한 트래픽 유발 등의 이유로 발생하는 크롤러의 차단을 최소화하고 능동적 병렬 방식의 크롤러를 구성함으로써, 웹 페이지 구성요소에 대한 분석 결과를 기반으로 필요 자원을 분배하여, 최적화된 크롤러 모듈의 병렬적인 실행을 통해 웹페이지 구성요소를 동시에 크롤링 할 수 있어 크롤링 과정에서 발생하는 소요 시간을 최소화할 수 있다.

[28]

도면의 간단한 설명

- [29] 도 1은 일 실시예에 있어서, 크롤러의 구조를 설명하기 위한 도면이다.
- [30] 도 2는 일 실시예에 있어서, 데이터 수집 방법을 설명하기 위한 흐름도이다.

- [31] 도 3은 일 실시예에 있어서, 탐색 크롤러 모듈의 동작을 설명하기 위한 흐름도이다.
- [32] 도 4는 일 실시예에 있어서, 스케줄러 모듈의 동작을 설명하기 위한 도면이다.
- [33] 도 5는 일 실시예에 있어서, 수집 크롤러 모듈의 동작을 설명하기 위한 도면이다.
- [34] 도 6은 일 실시예에 있어서, 다수 개의 노드로 구성된 서버 또는 클라우드 환경을 설명하기 위한 도면이다.

[35]

발명의 실시를 위한 최선의 형태

- [36] 이하, 실시예를 첨부한 도면을 참조하여 상세히 설명한다.

[37]

- [38] 도 1은 일 실시예에 있어서, 크롤러의 구조를 설명하기 위한 도면이다.

- [39] 데이터 수집 시스템(이하 '크롤러'로 기재하기로 함)(100)은 웹 페이지 내 크롤링에 요구되는 웹 페이지 정보를 탐색하는 탐색 크롤러 모듈(110), 탐색된 웹 페이지 정보에 기초하여 작업 공간을 동적으로 할당하는 스케줄러 모듈(120) 및 할당된 작업 공간에 기초하여 크롤링을 수행하는 수집 크롤러 모듈(130)로 구성될 수 있다.

- [40] 탐색 크롤러 모듈(101)은 웹 페이지의 구성 요소를 전체적으로 분석할 수 있다. 탐색 크롤러 모듈(101)은 DOM(Document Object Model)을 기반으로 웹 페이지(웹 사이트) 크롤링에 필요한 정보와 소요시간을 연산할 수 있다. 탐색 크롤러 모듈(101)은 DOM을 이용하여 웹 페이지 구성요소를 분석하고 구성요소와 관련된 링크에 리퀘스트(request)와 에코 패킷(echo packet)을 전송하여 콘텐츠의 크기(content_length)와 ping 명령어의 응답속도를 측정할 수 있다. 이후, 탐색 크롤러 모듈(101)은 전체 비율 1에서 ping 속도에 콘텐츠의 크기를 나눈 값을 감산하여 링크에 대한 가중치를 산정할 수 있다. 이때, 산정된 가중치는 스케줄러 모듈(102)에게 전달되며, 크롤러의 총 개수에 가중치를 곱해 해당 작업을 수행하는데 필요한 크롤러의 개수가 도출될 수 있다. 탐색 크롤러 모듈(101)은 웹 페이지를 탐색함에 따라 획득된 웹 페이지 정보를 공유 데이터 영역(104)에 저장할 수 있다.

- [41] 스케줄러 모듈(102)은 필요한 크롤러의 개수만큼 작업 공간(배열)을 동적 할당하고, 할당된 작업 공간이 가리키는 작업 정보(크롤러 할당 정보)를 공유 데이터(102)에 정의할 수 있다. 스케줄러 모듈(102)은 소요 시간을 종합하여 크롤링 모듈을 할당할 수 있다.

- [42] 수집 크롤러 모듈(103)은 실제 크롤링 기능을 담당할 수 있다. 수집 크롤러 모듈(130)은 작업 공간의 빈 공간을 찾아 들어가게 되고, 작업 공간에 있는 공유 데이터 영역(104)에 저장된 크롤링 진행상황을 불러와 이어서 진행할 작업이나 자신의 작업 구역을 확인한 후 크롤링 코드를 실행할 수 있다. 수집 크롤러 모듈(103)

은 크롤링을 실행함에 따라 수집된 정보를 공유 데이터 영역(104)에 저장할 수 있다.

- [43] 보다 상세하게는, 크롤러(100)는 사용자로부터 크롤링하기 위한 특정 웹 페이지에 대해 URL 정보(URL SEED)를 전달받아 웹 페이지 구성요소 및 하위 웹 페이지를 탐색하기 위한 탐색 크롤러 모듈(101), 탐색 크롤러 모듈(101)을 통해 수집된 웹 페이지 정보를 기반으로 수집 크롤러 모듈을 생성하기 위한 스케줄러 모듈(102), 스케줄러 모듈에 의해 할당받은 웹 페이지 구성요소의 정보를 크롤링하는 수집 크롤러 모듈(103), 탐색 크롤러 모듈(101), 스케줄러 모듈(102), 수집 크롤러 모듈(103)의 정보를 통합해서 관리하기 위한 공유 데이터 영역(104)으로 구성될 수 있다.
- [44] 탐색 크롤러 모듈(101)은 사용자에게 전달받은 URL 정보(즉, URL SEED)를 기반으로 하여 DOM 구조를 생성하고, 해당 웹 페이지에 리퀘스트와 에코 패킷을 전송하여 콘텐츠의 크기와 ping 명령어의 응답속도를 측정할 수 있다. 탐색 크롤러 모듈(101)은 측정된 응답속도에 기초하여 웹 페이지 구성요소의 크기 및 크롤링에 필요한 가중치를 연산할 수 있다. 탐색 크롤러 모듈(101)은 가중치 연산 결과를 스케줄러 모듈(102)에게 전달할 수 있다.
- [45] 스케줄러 모듈(102)은 탐색 크롤러 모듈(101)로부터 전달된 가중치를 기반으로 수집 크롤러 모듈(103)을 생성하기 위한 크롤러 풀을 활용함으로써, 크롤러 풀에서 가중치에 해당하는 크롤러의 개수를 할당할 수 있다.
- [46] 수집 크롤러 모듈(103)은 스케줄러 모듈(102)에 의해 웹 페이지 구성요소 및 하위 페이지 정보를 할당받아 크롤링 기능을 수행할 수 있다. 스케줄러 모듈(102)에 의해 생성된 크롤러 모듈(103)은 웹페이지 구성요소 및 하위 페이지의 개수에 따라 다수 개로 생성되어 병렬적으로 크롤링 연산을 수행할 수 있다. 수집 크롤러 모듈(103)의 각각은 통신 기능을 포함함으로써, 각 모듈 간의 크롤링 완료 여부 파악이 가능하므로, 수집 크롤러 모듈(103)에 의해 수집된 웹 페이지 정보는 공유 데이터 영역(104)로 전송될 수 있다. 정보 수집이 완료된 수집 크롤러 모듈(103)은 스케줄러 모듈(102)에게 완료 신호를 전송한 뒤 삭제되며 스케줄러 모듈(102)은 크롤러 풀에 새로운 크롤러 영역을 할당한다.
- [47] 도 2는 일 실시예에 있어서, 데이터 수집 방법을 설명하기 위한 흐름도이다.
- [48] 단계(210)에서 크롤러는 크롤링의 단일 기능을 수행하기 위해 모듈화된 다중 크롤러 모듈을 구성할 수 있다. 크롤러는 웹 페이지 내 크롤링에 요구되는 웹 페이지 정보를 탐색하는 탐색 크롤러 모듈, 탐색된 웹 페이지 정보에 기초하여 작업 공간을 동적으로 할당하는 스케줄러 모듈 및 할당된 작업 공간에 기초하여 크롤링 수행하는 수집 크롤러 모듈을 포함할 수 있다.
- [49] 단계(220)에서 크롤러는 구성된 다중 크롤러 모듈을 이용한 웹 페이지 접속을 통해 병렬적 크롤링을 수행할 수 있다. 크롤러는 탐색 크롤러 모듈을 통해 사용자에게 전달받은 URL 정보를 기반으로 DOM(Document Object Model) 구조를 생성하고, 리퀘스트(Request)와 에코(echo) 패킷을 전송함에 따라 콘텐츠의 크기

(content_length)와 ping 명령어의 응답속도를 측정할 수 있다. 크롤러는 측정된 응답속도에 기초하여 전체 비율 1에서 ping 명령어의 응답속도에 콘텐츠의 크기를 나눈 값을 감산하여 웹 페이지 구성요소의 크기 및 크롤링을 위한 가중치를 산정할 수 있다. 크롤러는 리퀘스트(Request)와 에코(echo) 패킷을 전송함에 따라 탐색된 웹 페이지 구성요소 및 하위 페이지를 포함하는 웹 페이지 정보를 공유 데이터 영역에 저장할 수 있다. 또한, 크롤러는 스케줄러 모듈을 통해 탐색 크롤러 모듈로부터 전달된 가중치를 기반으로 크롤러 풀을 이용하여 수집 크롤러 모듈의 개수를 연산하고, 연산된 수집 크롤러 모듈의 개수에 기초하여 수집 크롤러 모듈을 생성할 수 있다. 크롤러는 연산된 수집 크롤러 모듈의 개수에 기초하여 생성된 수집 크롤러 모듈을 작업 공간에 할당하고, 생성된 수집 크롤러 모듈에게 할당된 작업 공간이 가리키는 작업 정보를 전달할 수 있다. 또한, 크롤러는 수집 크롤러 모듈을 통해 스케줄러 모듈에 의해 할당된 작업 공간의 웹 페이지 구성요소 및 하위 페이지 정보를 포함하는 작업 정보를 통해 크롤링을 수행할 수 있다. 크롤러는 작업 공간에 연결되어 있는 진행 상황을 이용하여 이어서 진행할 작업이나 수집 크롤러 모듈의 작업 공간을 확인한 후, 크롤링을 수행할 수 있다. 크롤러는 수행된 크롤링이 완료되었는지 여부를 판단하고, 수행된 크롤링이 완료되지 않았을 경우 크롤링 진행 상황을 업데이트하고, 수행된 크롤링이 완료되었을 경우 스케줄러 모듈이 다음 작업 공간을 할당하였는지 여부를 판단할 수 있다. 크롤러는 스케줄러 모듈이 다음 작업 공간을 할당하지 않은 것으로 판단됨에 따라 크롤링 진행 상황을 업데이트하고 크롤러 대기열로 이동할 수 있다. 크롤러는 스케줄러 모듈이 다음 작업 공간을 할당한 것으로 판단됨에 따라 수집 크롤러 모듈을 할당된 다음 작업 공간에 배치할 수 있다. 크롤러는 수행된 크롤링을 통해 수집된 정보를 공유 데이터 영역에 저장할 수 있다.

[50] 도 3은 일 실시예에 있어서, 탐색 크롤러 모듈의 동작을 설명하기 위한 흐름도이다.

[51] 탐색 크롤러 모듈은 최초 웹 페이지를 탐색할 수 있다. 탐색 크롤러 모듈은 URL SEED를 통한 웹 페이지를 탐색할 수 있다(310). 탐색 크롤러 모듈은 탐색된 웹 페이지 내에 새로운 링크가 존재하는지 여부를 판단할 수 있다(320). 탐색 크롤러 모듈은 웹 페이지 구성요소 및 하위 웹 페이지에 접속이 가능한 새로운 링크가 존재한다면 해당 웹 페이지 구성요소를 대상으로 DOM을 제작할 수 있다(330). 탐색 크롤러 모듈은 웹 페이지 구성요소를 대상으로 DOM을 추가하여 구조를 분석한 뒤 예상 가중치를 계산하기 위해 리퀘스트와 에코 패킷을 전송함에 따라 콘텐츠의 크기와 ping 명령어의 응답속도를 측정하고, 측정된 응답속도에 기초하여 웹 페이지 구성요소의 크기 및 크롤링에 필요한 가중치를 계산할 수 있다(340). 탐색 크롤러 모듈은 가중치 계산이 완료됨에 따라 스케줄러 모듈에게 계산된 가중치 값을 전송할 수 있다(350). 이때, 스케줄러 모듈은 전송받은 가중치 값에 기초하여 크롤러 풀에서 수집 크롤러 모듈을 할당할 수 있다. 이와 같은 과정

으로 웹 페이지의 구조를 미리 파악하고 필요 자원을 계산하여 다른 크롤링 기법보다 빠르게 연속적으로 수집할 수 있게 된다.

[52] 도 4는 일 실시예에 있어서, 스케줄러 모듈의 동작을 설명하기 위한 도면이다.

[53] 스케줄러 모듈은 작업 공간의 공유 데이터에 크롤러의 행위를 저장할 수 있다(410). 스케줄러 모듈은 탐색 크롤러 모듈에서 전달받은 가중치 값을 이용하여 작업 공간을 동적으로 할당할 수 있다(420). 스케줄러 모듈은 크롤러 풀에서 가용할 수 있는 수집 크롤러 모듈의 개수를 연산하고, 연산된 수집 크롤러 모듈의 개수에 기초하여 수집 크롤러 모듈을 작업 공간에 동적으로 할당할 수 있다. 스케줄러 모듈은 크롤러 풀에 예약된 크롤러 모듈을 이용하여 새로운 수집 크롤러 모듈을 생성하고, 수집 크롤러 모듈의 생성이 완료되는 과정에서 탐색 크롤러 모듈에서 수집된 웹 페이지 구성요소 및 하위 페이지 정보를 할당할 수 있다. 이때, 스케줄러 모듈은 수집 크롤러 모듈 생성 과정에서 사용자의 요구 및 시스템 가용자원에 따라 각 웹 페이지 구성요소 및 하위 웹 페이지 개수와 동일한 수집 크롤러 모듈을 생성하거나, 웹 페이지 구성요소보다 적은 개수의 수집 크롤러 모듈을 생성할 수 있다. 스케줄러 모듈은 작업 공간에 남는 자리가 있는지 판단할 수 있다(430). 스케줄러 모듈은 작업 공간에 남는 자리가 없는 것으로 판단될 경우, 일정시간 대기할 수 있다. 스케줄러 모듈은 일정시간 대기 후, 다시 작업 공간에 남는 자리가 있는지 판단할 수 있다. 스케줄러 모듈은 작업 공간에 남는 자리가 있는 것으로 판단될 경우, 작업이 완료된 수집 크롤러 모듈이 존재하는지 판단할 수 있다(440). 스케줄러 모듈은 작업이 완료된 수집 크롤러 모듈이 존재하는 것으로 판단되는 경우, 남는 자리가 있는 작업 공간에 수집 크롤러 모듈을 배치할 수 있다(450). 스케줄러 모듈은 작업이 완료된 수집 크롤러 모듈이 존재하지 않는 것으로 판단되는 경우, 프로세스를 종료할 수 있다.

[54] 도 5는 일 실시예에 있어서, 수집 크롤러 모듈의 동작을 설명하기 위한 도면이다.

[55] 수집 크롤러 모듈은 스케줄러 모듈에 의해 생성될 수 있다. 수집 크롤러 모듈은 스케줄러 모듈 내의 작업 공간 할당을 대기할 수 있다(510). 수집 크롤러 모듈은 생성 과정에서 크롤링 대상 웹 페이지 구성요소 및 하위 페이지에 대한 정보를 할당받을 수 있다. 수집 크롤러 모듈은 할당받은 정보를 토대로 크롤링을 수행할 수 있다. 수집 크롤러 모듈은 작업 공간에 연결되어 있는 진행 상황과 코드를 가져와 이어서 크롤링을 실행할 수 있다(520). 수집 크롤러 모듈은 크롤링을 완료하였는지 여부를 판단할 수 있다(530). 수집 크롤러 모듈은 크롤링을 완료하지 않았을 경우, 진행상황을 업데이트할 수 있다(540). 수집 크롤러 모듈은 크롤링을 완료하였을 경우, 스케줄러 모듈이 다음 작업 공간을 할당하였는지 여부를 판단할 수 있다(550). 수집 크롤러 모듈은 스케줄러 모듈이 다음 작업을 할당한 것으로 판단됨에 따라 할당된 다음 작업 공간에 수집 크롤러 모듈을 배치할 수 있다(570). 수집 크롤러 모듈은 스케줄러 모듈이 다음 작업 공간을 할당하지 않은 것으로 판단됨에 따라 크롤링 진행 상황을 업데이트하고 크롤러 대기열로 이동할

수 있다(560). 이때, 수집 크롤러 모듈은 크롤링 수행 과정에서 실시간으로 웹 페이지 구성요소 및 웹 페이지의 크롤링 범위 및 현재 크롤링 진행 과정에 대한 정보를 공유 데이터 영역에 전송할 수 있다. 또한, 수집 크롤러 모듈은 수집 크롤링 모듈에 포함되어 있는 통신 모듈을 이용하여 다른 수집 크롤링 모듈 또는 공유 데이터 영역에 필요한 데이터를 요청하거나 전송할 수 있다.

[56] 도 6은 일 실시예에 있어서, 다수 개의 노드로 구성된 서버 또는 클라우드 환경을 설명하기 위한 도면이다.

[57] 크롤링 환경은 다수 개의 노드로 구성된 서버 또는 클라우드 환경에서 구동될 수 있다. 각 노드는 크롤링 모듈을 포함하며 각각 다른 네트워크 IP 등을 사용하는 환경으로 구성될 수 있다.

[58] 각 노드에는 각 크롤링 모듈을 구동하는데 필요한 개별 공유 데이터 영역이 구성되어 있으며, 각 노드에서 생성된 최종 데이터는 별도로 구성된 최상위 공유 데이터 영역으로 전송되어, 최종적으로 사용자는 최상위 공유 데이터 영역을 통해 크롤링 결과 및 크롤링 데이터를 확인할 수 있다.

[59]

발명의 실시를 위한 형태

[60] 이상에서 설명된 장치는 하드웨어 구성요소, 소프트웨어 구성요소, 및/또는 하드웨어 구성요소 및 소프트웨어 구성요소의 조합으로 구현될 수 있다. 예를 들어, 실시예들에서 설명된 장치 및 구성요소는, 예를 들어, 프로세서, 콘트롤러, ALU(arithmetic logic unit), 디지털 신호 프로세서(digital signal processor), 마이크로컴퓨터, FPGA(field programmable gate array), PLU(programmable logic unit), 마이크로프로세서, 또는 명령(instruction)을 실행하고 응답할 수 있는 다른 어떠한 장치와 같이, 하나 이상의 범용 컴퓨터 또는 특수 목적 컴퓨터를 이용하여 구현될 수 있다. 처리 장치는 운영 체제(OS) 및 상기 운영 체제 상에서 수행되는 하나 이상의 소프트웨어 애플리케이션을 수행할 수 있다. 또한, 처리 장치는 소프트웨어의 실행에 응답하여, 데이터를 접근, 저장, 조작, 처리 및 생성할 수도 있다. 이해의 편의를 위하여, 처리 장치는 하나가 사용되는 것으로 설명된 경우도 있지만, 해당 기술분야에서 통상의 지식을 가진 자는, 처리 장치가 복수 개의 처리 요소(processing element) 및/또는 복수 유형의 처리 요소를 포함할 수 있음을 알 수 있다. 예를 들어, 처리 장치는 복수 개의 프로세서 또는 하나의 프로세서 및 하나의 콘트롤러를 포함할 수 있다. 또한, 병렬 프로세서(parallel processor)와 같은, 다른 처리 구성(processing configuration)도 가능하다.

[61] 소프트웨어는 컴퓨터 프로그램(computer program), 코드(code), 명령(instruction), 또는 이들 중 하나 이상의 조합을 포함할 수 있으며, 원하는 대로 동작하도록 처리 장치를 구성하거나 독립적으로 또는 결합적으로(collectively) 처리 장치를 명령할 수 있다. 소프트웨어 및/또는 데이터는, 처리 장치에 의하여 해석되거나 처리 장치에 명령 또는 데이터를 제공하기 위하여, 어떤 유형의 기계,

구성요소(component), 물리적 장치, 가상 장치(virtual equipment), 컴퓨터 저장 매체 또는 장치에 구체화(embodiment)될 수 있다. 소프트웨어는 네트워크로 연결된 컴퓨터 시스템 상에 분산되어서, 분산된 방법으로 저장되거나 실행될 수도 있다. 소프트웨어 및 데이터는 하나 이상의 컴퓨터 판독 가능 기록 매체에 저장될 수 있다.

- [62] 실시예에 따른 방법은 다양한 컴퓨터 수단을 통하여 수행될 수 있는 프로그램 명령 형태로 구현되어 컴퓨터 판독 가능 매체에 기록될 수 있다. 상기 컴퓨터 판독 가능 매체는 프로그램 명령, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 상기 매체에 기록되는 프로그램 명령은 실시예를 위하여 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 당업자에게 공지되어 사용 가능한 것일 수도 있다. 컴퓨터 판독 가능 기록 매체의 예에는 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체(magnetic media), CD-ROM, DVD와 같은 광기록 매체(optical media), 플롭티컬 디스크(floptical disk)와 같은 자기-광 매체(magneto-optical media), 및 롬(ROM), 램(RAM), 플래시 메모리 등과 같은 프로그램 명령을 저장하고 수행하도록 특별히 구성된 하드웨어 장치가 포함된다. 프로그램 명령의 예에는 컴파일러에 의해 만들어지는 것과 같은 기계어 코드 뿐만 아니라 인터프리터 등을 사용해서 컴퓨터에 의해서 실행될 수 있는 고급 언어 코드를 포함한다.
- [63] 이상과 같이 실시예들이 비록 한정된 실시예와 도면에 의해 설명되었으나, 해당 기술분야에서 통상의 지식을 가진 자라면 상기의 기재로부터 다양한 수정 및 변형이 가능하다. 예를 들어, 설명된 기술들이 설명된 방법과 다른 순서로 수행되거나, 및/또는 설명된 시스템, 구조, 장치, 회로 등의 구성요소들이 설명된 방법과 다른 형태로 결합 또는 조합되거나, 다른 구성요소 또는 균등물에 의하여 대치되거나 치환되더라도 적절한 결과가 달성될 수 있다.
- [64] 그러므로, 다른 구현들, 다른 실시예들 및 특허청구범위와 균등한 것들도 후술하는 특허청구범위의 범위에 속한다.

청구범위

- [청구항 1] 데이터 수집 시스템에 의해 수행되는 데이터 수집 방법에 있어서,
크롤링의 단일 기능을 수행하기 위해 모듈화된 다중 크롤러 모듈을 구성하는 단계; 및
상기 구성된 다중 크롤러 모듈을 이용한 웹 페이지 접속을 통해 병렬적 크롤링을 수행하는 단계를 포함하는 데이터 수집 방법.
- [청구항 2] 제1항에 있어서,
상기 다중 크롤러 모듈은,
웹 페이지 내 크롤링에 요구되는 웹 페이지 정보를 탐색하는 탐색 크롤러 모듈, 상기 탐색된 웹 페이지 정보에 기초하여 작업 공간을 동적으로 할당하는 스케줄러 모듈 및 상기 할당된 작업 공간에 기초하여 크롤링 수행하는 수집 크롤러 모듈을 포함하는 것을 특징으로 하는 데이터 수집 방법.
- [청구항 3] 제1항에 있어서,
상기 병렬적 크롤링을 수행하는 단계는,
탐색 크롤러 모듈에서, 사용자에게 전달받은 URL 정보를 기반으로 DOM(Document Object Model) 구조를 생성하고, 리퀘스트(Request)와 에코(echo) 패킷을 전송함에 따라 콘텐츠의 크기(content_length)와 ping 명령어의 응답속도를 측정하는 단계를 포함하는 데이터 수집 방법.
- [청구항 4] 제3항에 있어서,
상기 병렬적 크롤링을 수행하는 단계는,
상기 측정된 응답속도에 기초하여 전체 비율 1에서 상기 ping 명령어의 응답속도에 콘텐츠의 크기를 나눈 값을 감산하여 웹 페이지 구성요소의 크기 및 크롤링을 위한 가중치를 산정하는 단계를 포함하는 데이터 수집 방법.
- [청구항 5] 제3항에 있어서,
상기 병렬적 크롤링을 수행하는 단계는,
상기 리퀘스트(Request)와 에코(echo) 패킷을 전송함에 따라 탐색된 웹 페이지 구성요소 및 하위 페이지를 포함하는 웹 페이지 정보를 공유 데이터 영역에 저장하는 단계를 포함하는 데이터 수집 방법.
- [청구항 6] 제1항에 있어서,
상기 병렬적 크롤링을 수행하는 단계는,
스케줄러 모듈에서, 탐색 크롤러 모듈로부터 전달된 가중치를 기반으로 크롤러 풀을 이용하여 수집 크롤러 모듈의 개수를 연산하고, 상기 연산된 수집 크롤러 모듈의 개수에 기초하여 수집 크롤러 모듈을 생성하는 단계

- 를 포함하는 데이터 수집 방법.
- [청구항 7] 제6항에 있어서,
상기 병렬적 크롤링을 수행하는 단계는,
상기 연산된 수집 크롤러 모듈의 개수에 기초하여 생성된 수집 크롤러 모듈을 작업 공간에 할당하고, 상기 생성된 수집 크롤러 모듈에게 상기 할당된 작업 공간이 가리키는 작업 정보를 전달하는 단계
를 포함하는 데이터 수집 방법.
- [청구항 8] 제1항에 있어서,
상기 병렬적 크롤링을 수행하는 단계는,
수집 크롤러 모듈에서, 스케줄러 모듈에 의해 할당된 작업 공간의 웹 페이지 구성요소 및 하위 페이지 정보를 포함하는 작업 정보를 통해 크롤링을 수행하는 단계
를 포함하는 데이터 수집 방법.
- [청구항 9] 제8항에 있어서,
상기 병렬적 크롤링을 수행하는 단계는,
상기 작업 공간에 연결되어 있는 진행 상황을 이용하여 이어서 진행할 작업이나 수집 크롤러 모듈의 작업 공간을 확인한 후, 크롤링을 수행하는 단계
를 포함하는 데이터 수집 방법.
- [청구항 10] 제9항에 있어서,
상기 병렬적 크롤링을 수행하는 단계는,
상기 수행된 크롤링이 완료되었는지 여부를 판단하고, 상기 수행된 크롤링이 완료되지 않았을 경우 크롤링 진행 상황을 업데이트하고, 상기 수행된 크롤링이 완료되었을 경우 스케줄러 모듈이 다음 작업 공간을 할당하였는지 여부를 판단하는 단계
를 포함하는 데이터 수집 방법.
- [청구항 11] 제10항에 있어서,
상기 병렬적 크롤링을 수행하는 단계는,
상기 스케줄러 모듈이 다음 작업 공간을 할당하지 않은 것으로 판단됨에 따라 크롤링 진행 상황을 업데이트하고 크롤러 대기열로 이동하는 단계
를 포함하는 데이터 수집 방법.
- [청구항 12] 제11항에 있어서,
상기 병렬적 크롤링을 수행하는 단계는,
상기 스케줄러 모듈이 다음 작업 공간을 할당한 것으로 판단됨에 따라 수집 크롤러 모듈을 할당된 다음 작업 공간에 배치하는 단계
를 포함하는 데이터 수집 방법.
- [청구항 13] 제9항에 있어서,
상기 병렬적 크롤링을 수행하는 단계는,

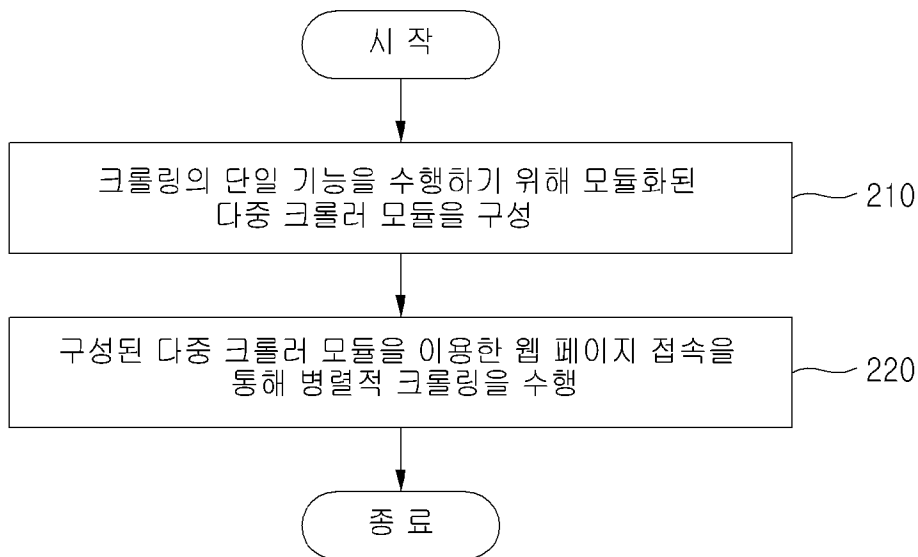
상기 수행된 크롤링을 통해 수집된 정보를 공유 데이터 영역에 저장하는 단계

를 포함하는 데이터 수집 방법.

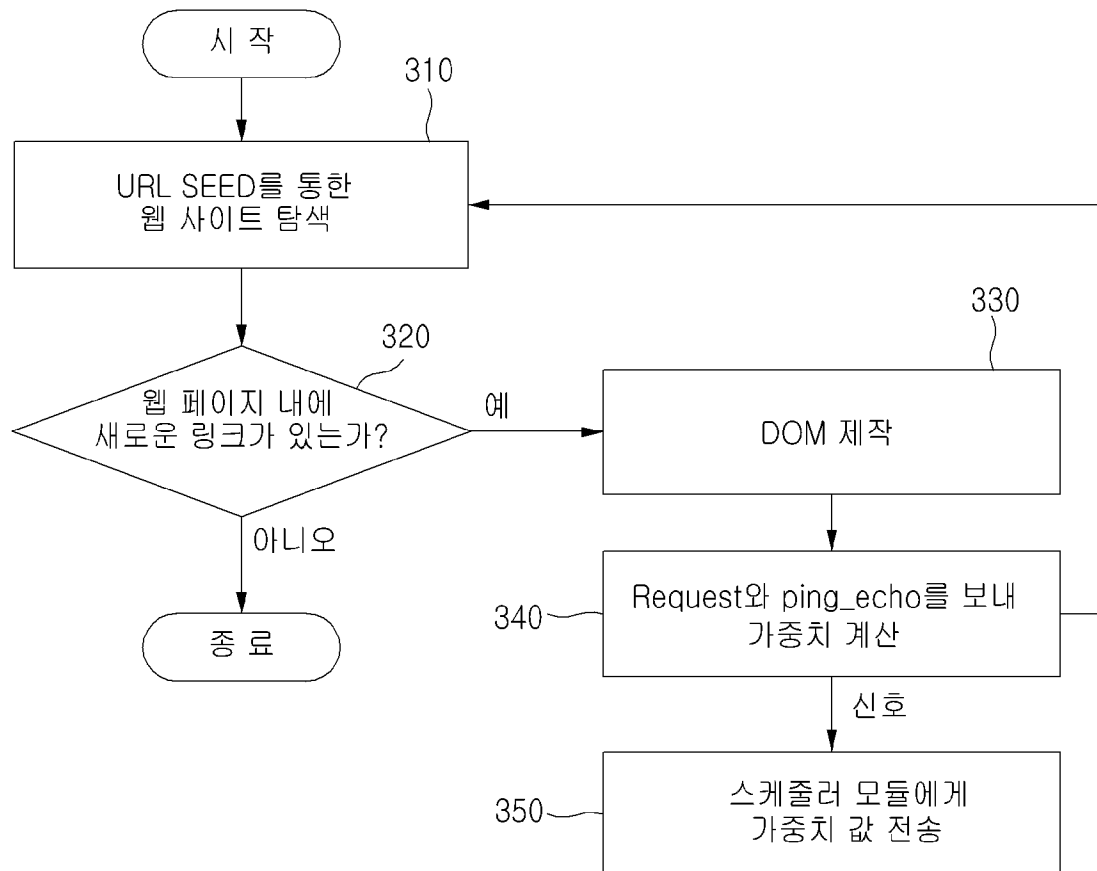
[청구항 14] 제1항 내지 제13항 중 어느 한 항의 데이터 수집 방법을 상기 데이터 수집 시스템에 실행시키기 위해 비-일시적인 컴퓨터 판독가능한 기록 매체에 저장되는 컴퓨터 프로그램.

[청구항 15] 데이터 수집 시스템에 있어서,
메모리에 포함된 컴퓨터 판독가능한 명령들을 실행하도록 구성된 적어도 하나의 프로세서
를 포함하고,
상기 적어도 하나의 프로세서는,
크롤링의 단일 기능을 수행하기 위해 모듈화된 다중 크롤러 모듈을 구성하고,
상기 구성된 다중 크롤러 모듈을 이용한 웹 페이지 접속을 통해 병렬적 크롤링을 수행하는
것을 특징으로 하는 데이터 수집 시스템.

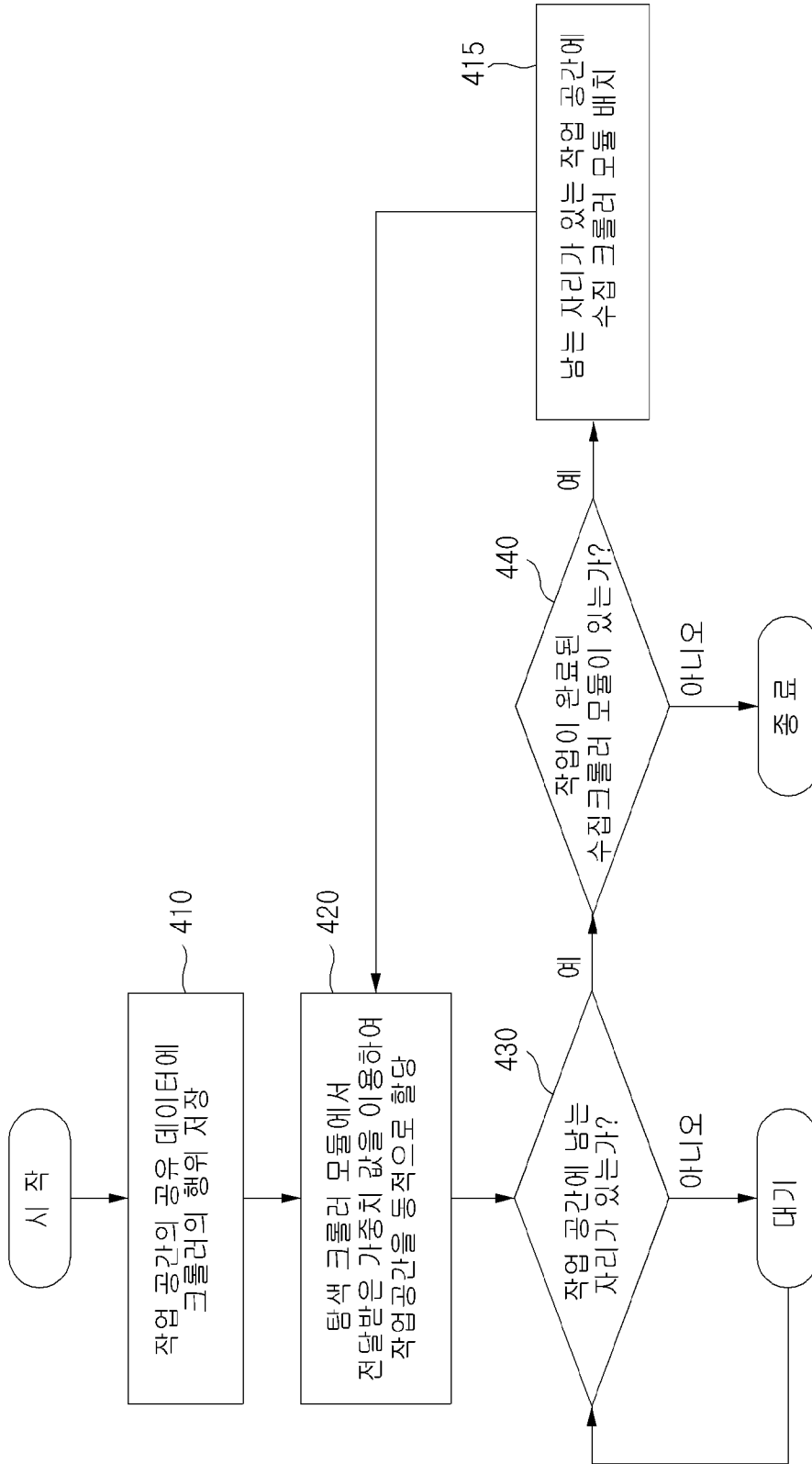
[도2]



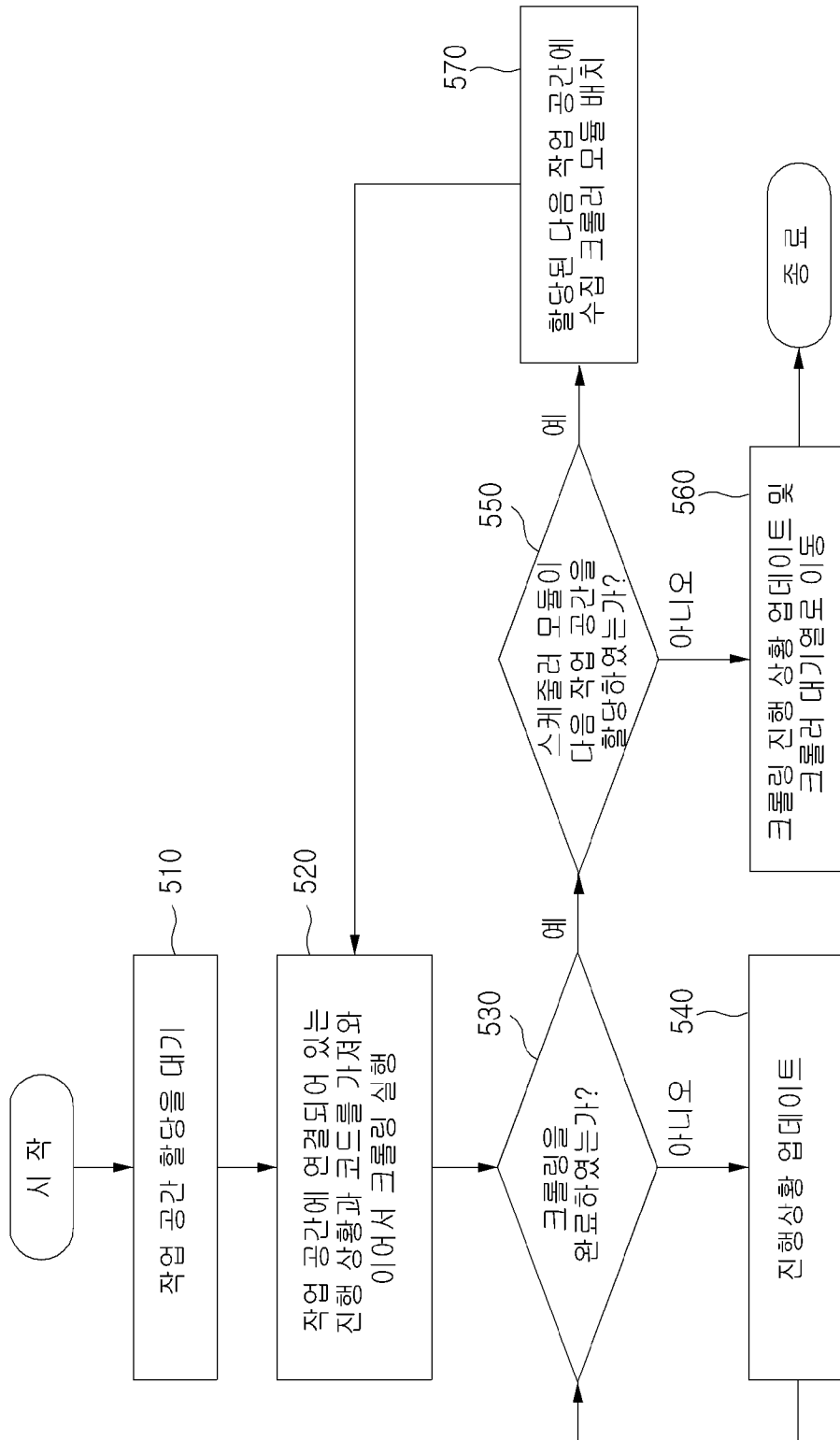
[도3]



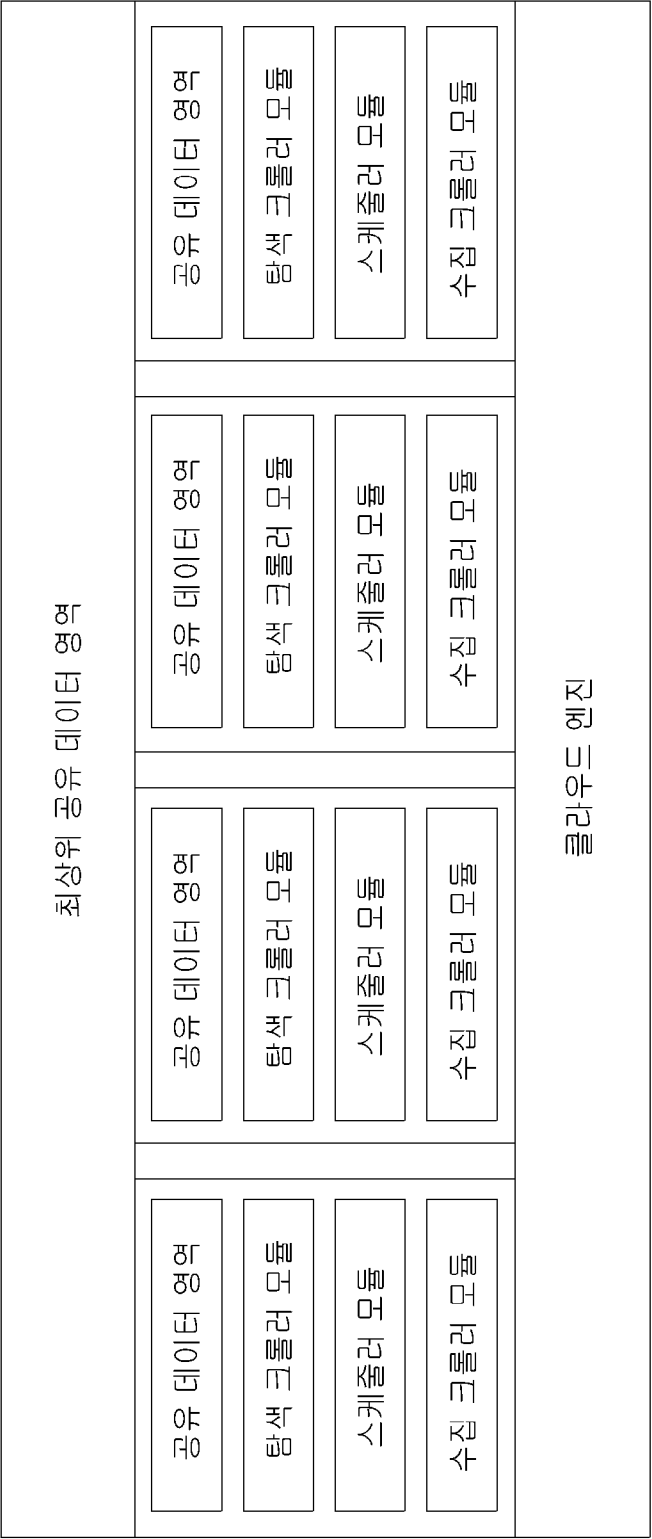
[도4]



[도5]



[도6]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/KR2024/013922

A. CLASSIFICATION OF SUBJECT MATTER G06F 16/951(2019.01)i; G06F 16/955(2019.01)i; G06F 9/48(2006.01)i; G06F 9/50(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) G06F 16/951(2019.01); G06F 15/173(2006.01); G06F 16/00(2019.01); G06F 16/901(2019.01); G06F 17/30(2006.01); G06F 9/54(2006.01) 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: 다중 크롤러(multiple crawlers), 웹페이지(web-pages), 병렬(parallel), 스케줄러(scheduler), 할당(allocation)		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 7546370 B1 (ACHARYA, Anurag et al.) 09 June 2009 (2009-06-09) See column 2, line 67 - column 3, line 3 and column 6, lines 35-37; and figures 1 and 3.	1,14-15
Y		2-13
Y	US 2012-0109927 A1 (PRASAD, Mukul Ranjan) 03 May 2012 (2012-05-03) See paragraph [0039]; and figure 1.	2,9-13
Y	KR 10-2442535 B1 (KN LAB, INC. et al.) 13 September 2022 (2022-09-13) See paragraphs [0030] and [0034]; claim 1; and figure 2.	3-13
X	KR 10-2043508 B1 (THE INDUSTRY & ACADEMIC COOPERATION IN CHUNGNAM NATIONAL UNIVERSITY (IAC)) 11 November 2019 (2019-11-11) See paragraphs [0031] and [0050]; and figure 1.	1,14-15
☒ 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 “D” document cited by the applicant in the international application “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 20 December 2024		Date of mailing of the international search report 20 December 2024
Name and mailing address of the ISA/KR Korean Intellectual Property Office Government Complex-Daejeon Building 4, 189 Cheongsaro, Seo-gu, Daejeon 35208 Facsimile No. +82-42-481-8578		Authorized officer Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/KR2024/013922

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2015-0286725 A1 (ZILLABYTE, INC.) 08 October 2015 (2015-10-08) See paragraphs [0055]-[0070]; claims 3-4; and figure 1.	1-15

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/KR2024/013922

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)		Publication date (day/month/year)
US	7546370	B1	09 June 2009	None		
US	2012-0109927	A1	03 May 2012	EP	2447893 A2	02 May 2012
				EP	2447898 A3	31 October 2012
				JP	2012-099102 A	24 May 2012
				JP	5742666 B2	01 July 2015
				US	2012-0109931 A1	03 May 2012
				US	2012-0110063 A1	03 May 2012
				US	8832065 B2	09 September 2014
KR	10-2442535	B1	13 September 2022	None		
KR	10-2043508	B1	11 November 2019	KR	10-2019-0055366 A	23 May 2019
US	2015-0286725	A1	08 October 2015	None		

A. 발명이 속하는 기술분류(국제특허분류(IPC))

G06F 16/951(2019.01)i; G06F 16/955(2019.01)i; G06F 9/48(2006.01)i; G06F 9/50(2006.01)i

B. 조사된 분야

조사된 최소문헌(국제특허분류를 기재)

G06F 16/951(2019.01); G06F 15/173(2006.01); G06F 16/00(2019.01); G06F 16/901(2019.01); G06F 17/30(2006.01); G06F 9/54(2006.01)

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: 다중 크롤러(multiple crawlers), 웹 페이지(web-pages), 병렬(parallel), 스케줄러(scheduler), 할당(allocation)

C. 관련 문헌

카테고리*	인용문헌명 및 관련 구절(해당하는 경우)의 기재	관련 청구항
X	US 7546370 B1 (ANURAG ACHARYA 등) 2009.06.09 컬럼 2, 라인 67 - 컬럼 3, 라인 3; 컬럼 6, 라인 35-37; 및 도면 1, 3	1,14-15
Y		2-13
Y	US 2012-0109927 A1 (MUKUL RANJAN PRASAD) 2012.05.03 단락 [0039]; 및 도면 1	2,9-13
Y	KR 10-2442535 B1 ((주)케이엔랩 등) 2022.09.13 단락 [0030], [0034]; 청구항 1; 및 도면 2	3-13
X	KR 10-2043508 B1 (충남대학교산학협력단) 2019.11.11 단락 [0031], [0050]; 및 도면 1	1,14-15
A	US 2015-0286725 A1 (ZILLABYTE, INC.) 2015.10.08 단락 [0055]-[0070]; 청구항 3-4; 및 도면 1	1-15

☐ 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

“D” 본 국제출원에서 출원인이 인용한 문헌

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

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

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

“P” 우선일 이후에 공개되었으나 국제출원일 이전에 공개된 문헌

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

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

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

“&” 동일한 대응특허문헌에 속하는 문헌

국제조사의 실제 완료일

2024년12월20일(20.12.2024)

국제조사보고서 발송일

2024년12월20일(20.12.2024)

ISA/KR의 명칭 및 우편주소

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

팩스 번호 +82-42-481-8578

심사관

양정록

전화번호 +82-42-481-5709

국제조사보고서에서 인용된 특허문헌	공개일	대응특허문헌	공개일
US 7546370 B1	2009/06/09	없음	
US 2012-0109927 A1	2012/05/03	EP 2447893 A2	2012/05/02
		EP 2447898 A3	2012/10/31
		JP 2012-099102 A	2012/05/24
		JP 5742666 B2	2015/07/01
		US 2012-0109931 A1	2012/05/03
		US 2012-0110063 A1	2012/05/03
		US 8832065 B2	2014/09/09
KR 10-2442535 B1	2022/09/13	없음	
KR 10-2043508 B1	2019/11/11	KR 10-2019-0055366 A	2019/05/23
US 2015-0286725 A1	2015/10/08	없음	