



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2014-0069756
(43) 공개일자 2014년06월10일

(51) 국제특허분류(Int. Cl.)
G06F 17/40 (2006.01) G06F 15/16 (2006.01)
G06F 9/06 (2006.01)
(21) 출원번호 10-2012-0137439
(22) 출원일자 2012년11월29일
심사청구일자 2012년11월29일

(71) 출원인
대한민국(국가기록원)
대전광역시 서구 청사로 189, 2동 406호 (둔산동, 정부대전청사)
(72) 발명자
차주용
대전 유성구XXXXXXXXXXXX (신성동, XXXXXXXXXXXXX)
이규철
대전 유성구XXXXXXXXXXXXXXXXXXXX (용산동, XXXXXXXXXXXXXXX)
(74) 대리인
맹성재

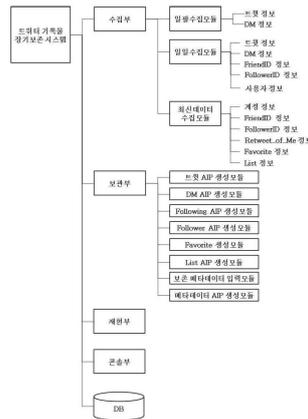
전체 청구항 수 : 총 18 항

(54) 발명의 명칭 트위터 기록물 장기보존 시스템

(57) 요약

본 발명은 행정기관에서 운영하고 있는 트위터 기록물을 대상으로 장기 보존하고 재현 가능한 메타데이터 및 포맷을 제공하는 시스템에 관한 것으로서, 트위터 기록물을 수집하는 수집부, 상기 수집부에서 수집한 정보를 구조적 메타데이터 문서로 생성하는 보관부, 상기 보관부에서 생성된 메타데이터를 트윗 문서로 재현하는 재현부, 상기 수집부, 보관부 및 재현부를 모니터링하고 제어하는 콘솔부 및 상기 수집부에서 수집한 데이터를 저장하는 DB부를 포함하는 것을 특징으로 한다.

대표도 - 도2



이 발명을 지원한 국가연구개발사업

과제고유번호 2012-0817

부처명 행정안전부

연구사업명 복합 전자기록물 장기보존 기반기술 연구

연구과제명 행정기관의 SNS 기록물 수집·보존·재현 도구 연구개발

기여율 1/1

주관기관 충남대학교

연구기간 2012.05.02 ~ 2012.11.30

특허청구의 범위

청구항 1

트위터 기록물을 수집하는 수집부;

상기 수집부에서 수집한 정보를 구조적 메타데이터 문서로 생성하는 보관부;

상기 보관부에서 생성된 메타데이터를 트윗 문서로 재현하는 재현부;

상기 수집부, 보관부 및 재현부를 모니터링하고 제어하는 콘솔부; 및

상기 수집부에서 수집한 데이터를 저장하는 DB부를 포함하는 것을 특징으로 하는 트위터 기록물 장기보존 시스템.

청구항 2

제1항에 있어서,

상기 수집부는 일괄수집모듈, 일일수집모듈 및 최신데이터수집모듈로 구성된 것을 특징으로 하는 트위터 기록물 장기보존 시스템.

청구항 3

제2항에 있어서,

상기 일괄수집모듈은 행정기관에서 등록한 예전 자료를 최초로 수집하는 모듈이고, 트윗(tweet) 정보는 트위터사의 서버에 현재부터 보관된 최대건수를 일괄적으로 수집하고, DM(direct message) 정보는 현재부터 보관된 최대건수를 일괄적으로 수집하는 것을 특징으로 하는 트위터 기록물 장기보존 시스템.

청구항 4

제2항에 있어서,

상기 일일수집모듈은 행정기관에서 등록한 최근의 트윗과 DM 정보를 일일로 수집하고, 팔로잉하고 있는 사용자 정보인 FriendID, 팔로워 하고 있는 사용자 정보인 FollowerID 및 사용자정보를 수집하여 저장하는 것을 특징으로 하는 트위터 기록물 장기보존 시스템.

청구항 5

제4항에 있어서,

일일수집 대상이 되는 자료는 스케줄러(cron)에 의하여 일별 및 월별로 주기적으로 수집하는 것을 특징으로 하는 트위터 기록물 장기보존 시스템.

청구항 6

제2항에 있어서,

상기 최신데이터수집모듈은 아카이빙(archiving) 직전의 최신 정보를 일회 수집하는 모듈이고, 행정기관의 계정 정보, FriendID, FollowerID, 트윗을 리트윗하는 사용자 정보인 Retweet_of_Me, 관심글로 등록한 트윗 정보인 Favorite 및 그룹별 관리를 위해 등록한 List를 수집하는 것을 특징으로 하는 트위터 기록물 장기보존 시스템.

청구항 7

제1항에 있어서,

상기 보관부는 상기 수집부에서 수집한 정보를 영구적 보존 파일로 만들기 위하여 xml 데이터를 생성하는 하는 것을 특징으로 하는 트위터 기록물 장기보존 시스템.

청구항 8

제7항에 있어서,

상기 보관부는 트윗 AIP(Archive Information Package) 생성모듈, DM AIP 생성모듈, Following AIP 생성모듈, Follower AIP 생성모듈, Favorite 생성모듈, List 생성모듈, 보존 메타데이터 입력모듈 및 메타데이터 AIP 생성모듈로 구성된 것을 특징으로 하는 트위터 기록물 장기보존 시스템.

청구항 9

제8항에 있어서,

상기 트윗 AIP 생성모듈은 상기 수집부에서 수집한 트윗 정보와 리트윗 정보를 바탕으로 트윗 AIP를 생성하여 xml 데이터를 생성하는 것을 특징으로 하는 트위터 기록물 장기보존 시스템.

청구항 10

제8항에 있어서,

상기 DM AIP 생성모듈은 상기 수집부에서 수집한 DM 정보를 바탕으로 DM AIP를 생성하여 xml 데이터를 생성하는 것을 특징으로 하는 트위터 기록물 장기보존 시스템.

청구항 11

제8항에 있어서,

상기 Favorite AIP 생성모듈은 상기 수집부에서 수집한 Favorite 정보를 바탕으로 Favorite AIP를 생성하여 xml 데이터를 생성하는 것을 특징으로 하는 트위터 기록물 장기보존 시스템.

청구항 12

제8항에 있어서,

상기 Follower AIP 생성모듈은 상기 수집부에서 수집한 Follower 정보를 바탕으로 Follower AIP를 생성하는 것을 특징으로 하는 트위터 기록물 장기보존 시스템.

청구항 13

제8항에 있어서,

상기 Friend AIP 생성모듈의 상기 수집부에서 수집한 Friend 정보를 바탕으로 Friend AIP를 생성하여 xml 데이터를 생성하는 것을 특징으로 하는 트위터 기록물 장기보존 시스템.

청구항 14

제8항에 있어서,

상기 List AIP 생성모듈의 상기 수집부에서 수집한 List 정보를 바탕으로 List AIP를 생성하여 xml 데이터를 생성하는 것을 특징으로 하는 트위터 기록물 장기보존 시스템.

청구항 15

제8항에 있어서,

상기 보존 메타데이터 입력모듈은 메타데이터 AIP 생성을 위한 정보를 입력하는 것을 특징으로 하는 트위터 기록물 장기보존 시스템.

청구항 16

제8항에 있어서,

상기 메타데이터 AIP 생성모듈은 수집 및 입력한 정보를 바탕으로 메타데이터 AIP를 생성하는 것을 특징으로 하는 트위터 기록물 장기보존 시스템.

청구항 17

제1항 내지 제16항 중 어느 한 항의 트위터 기록물 장기보존 시스템을 이용한 트위터 기록물의 장기보존 포맷에 있어서,

상기 포맷은 연단위의 기록물철과 기록물건으로 구성된 것을 특징으로 하는 트위터 기록물의 장기보존 포맷.

청구항 18

제17항에 있어서,

상기 기록물건은 메타데이터 문서이고, 트윗, DM, Friends, Followers, Lists, Favorites, 통계문서(status)로 구성된 것을 특징으로 하는 트위터 기록물의 장기보존 포맷.

명세서

기술분야

[0001] 본 발명은 트위터 기록물 장기보존을 위한 시스템에 관한 것으로서, 상세하게는 우리나라 모든 행정기관에서 운영하고 있는 트위터 기록물을 대상으로 장기 보존하고 재현 가능한 메타데이터 및 포맷을 제공하는 시스템에 관한 것이다.

배경기술

[0002] 2006년 미국의 벤처기업 오비어스(Obvious)사가 처음 개설한 트위터(twitter)는 블로그의 인터페이스에 미니홈페이지의 친구맺기 기능, 메신저의 신속성을 갖춘 소셜 네트워크 서비스(Social network Service: SNS)로서, 젊은 세대들만의 교류 수단으로 사용하던 기존 서비스와는 달리 국내외 저명인사들도 다수의 팔로워(follower)를 보유하면서 일반 대중과 소통하는데 이용하는 등 사용자 범위가 다양화되고 정보의 빠른 공유와 확산력으로 전 세계적 트렌드로 자리잡아, 선거나 기업들의 홍보 등에 활용되고 있으며 우리나라의 행정기관에서도 정책을 홍보하고 국민들의 의견을 수렴하기에 이르렀다.

[0003] 국민들의 의견 수렴 및 각 행정기관의 정책을 국민들에게 홍보하기 위한 목적으로 사용되는 각 행정기관의 트위터 기록물은 미래에 역사적 기록 및 법적 증거로 활용될 수 있다. 그러나 SNS 제공업체가 영구히 서비스를 제공한다는 보장이 없어 서비스를 중단하거나, 현재의 트위터사처럼 정책에 따라 최대 보관건수(현재부터 3,200건)를 제한하는 경우 수많은 과거의 기록은 사라지게 되므로 SNS 기록물들을 장기적으로 보존할 필요성이 대두되었다.

[0004] 미국의 경우 국가정보 인프라 스트럭처 및 보존계획(NDIIPP)를 중심으로 트위터 기록물에 대한 아카이빙(achieving) 프로젝트가 진행되고 있으며, 미국 아카비스트협회(SAA)를 통해 트위터 기록물 아카이빙 방법론을 연구 중에 있다. 국내의 경우 국가기록원을 중심으로 SNS 기록물 아카이빙 방법을 연구하고 있으나 현재는 SNS 데이터를 백업하는 초기 단계이며 보다 발전된 체계적인 아카이빙 방법이 필요하다.

[0005] 주요 선행기술로는 특허문헌 1에 제시된 트윗 북 서비스 제공시스템으로서, SNS인 트위터에 자신이 남긴 글인 트윗(tweet)과 타인의 글을 자신의 트위터로 옮겨온 리트윗(retweet) 등의 글을 추출하여 자신만의 책으로 소유할 수 있도록 다양한 조건에 의하여 추출하고 다양한 템플릿을 이용하여 레이아웃을 형성하여 책으로 제공하는 것을 특징으로 한다. 이러한 종래의 기술은 자신이 남긴 기록물을 정리하여 책으로 만드는 것으로서 개인 기록물에 국한되고 개인 취향에 초점이 맞추어지기 때문에 국가적 기록물로 영구보존을 위한 체계적 방법으로는 부족한 점이 있다.

선행기술문헌

특허문헌

[0006] (특허문헌 0001) 1. 한국공개특허 제2012-0046806호(2012. 5. 11. 공개)

발명의 내용

해결하려는 과제

[0007] 상기의 필요성에 따라 본 발명에 따른 발명은, 우리나라 모든 행정기관에서 운영하고 있는 트위터 기록물을 대상으로 장기 보존하는데 필요한 메타데이터 및 포맷을 생성하는 시스템을 제공하고자 한다.

[0008] 또한, ISO 표준(ISO 14271)인 OAIS 참조모형을 따르고 있는 기존 전자기록물과 웹 기록물 장기보존 메타데이터 및 호환을 고려하고, 재현 가능한 포맷 및 메타데이터를 제공하고자 한다.

과제의 해결 수단

[0009] 상기의 해결하고자 하는 과제의 해결 수단으로서 트위터 기록물 장기보존 시스템은, 트위터 기록물을 수집하는 수집부, 상기 수집부에서 수집한 정보를 구조적 메타데이터 문서로 생성하는 보관부, 상기 보관부에서 생성된 메타데이터를 트윗 문서로 재현하는 재현부, 상기 수집부, 보관부 및 재현부를 모니터하고 제어하는 콘솔부 및 상기 수집부에서 수집한 데이터를 저장하는 DB부를 포함하는 것을 특징으로 한다.

[0010] 본 발명의 바람직한 실시예로서, 상기 수집부는 일괄수집모듈, 일일수집모듈 및 최신데이터수집모듈로 구성되고, 상기 일괄수집모듈은 행정기관에서 등록된 예전 자료를 최초로 수집하는 모듈이고, 트윗 정보는 트위터사의 서버에 보관된 현재부터 3,200건의 예전 자료를 일괄적으로 수집하고, DM 정보는 현재부터 800건의 예전 자료를 일괄적으로 수집하는 것을 특징으로 한다.

[0011] 본 발명의 바람직한 실시예로서, 상기 일일수집모듈은 행정기관에서 등록된 최근의 트윗과 DM을 일일로 수집하고, 팔로잉하고 있는 사용자 정보인 FriendID, 팔로워 하고 있는 사용자 정보인 FollowerID 및 사용자정보를 수집하고, 일일수집 대상이 되는 자료는 스케줄러(cron)에 의하여 일별 및 월별로 주기적으로 수집하는 것을 특징으로 한다.

[0012] 본 발명의 바람직한 실시예로서, 상기 최신데이터수집모듈은 아카이빙 직전의 최신 정보를 일회 수집하는 모듈이고, 행정기관의 계정정보, FriendID, FollowerID, 트윗을 리트윗하는 사용자 정보인 Retweet_of_Me, 관심글로 등록된 트윗 정보인 Favorite 및 그룹별 관리를 위해 등록된 List를 수집하는 것을 특징으로 한다.

[0013] 본 발명의 바람직한 실시예로서, 상기 보관부는 상기 수집부에서 수집한 정보를 영구적 보존 파일로 만들기 위하여 xml 생성하는 하는 기능을 하고, 트윗 AIP(Archive Information Package) 생성모듈, DM AIP 생성모듈, Following AIP 생성모듈, Follower AIP 생성모듈, Favorite 생성모듈, List 생성모듈, 보존 메타데이터 입력모듈 및 메타데이터 AIP 생성모듈로 구성된 것을 특징으로 한다.

[0014] 본 발명의 다른 실시예로서, 트위터 기록물 장기보존 시스템을 이용한 트위터 기록물의 장기보존 포맷은 연단위의 기록물철과 기록물건으로 구성되고, 상기 기록물건은 메타데이터 문서이고, 트윗, DM, Friends, Followers, Lists, Favorites, 통계문서(status)로 구성된 것을 특징으로 한다.

발명의 효과

[0015] 현재의 트위터사처럼 정책에 따라 최대 보관건수(현재부터 3,200건)를 제한하는 경우 수많은 과거의 기록은 사라지게 되나, 본 발명에 의한 트위터 기록물 장기보존을 위한 시스템을 사용하게 되면, 국민들의 의견 수렴 및 각 행정기관의 정책을 국민들에게 홍보된 트위터 기록물은 미래에 역사적 기록 및 법적 증거로 활용될 수 있다.

[0016] 또한, 본 발명에 의한 트위터 기록물은 ISO 표준(ISO 14271)인 OAIS 참조모형을 따르고 있는 기존 전자기록물과 웹 기록물 장기보존 메타데이터와 호환이 가능하고, 다시 트윗 정보 형태로 재현이 가능하다.

도면의 간단한 설명

- [0017] 도 1은 본 발명에서의 데이터 수집 절차도
- 도 2는 본 발명에 따른 트위터 기록물 장기보존 시스템
- 도 3은 트윗 정보에 대한 일괄수집모듈의 알고리즘 순서도
- 도 4는 트윗 정보에 대한 일괄수집모듈을 화면상에 구현한 프로그램을 캡처한 그림
- 도 5는 트윗 정보에 대한 일일수집모듈의 알고리즘 순서도
- 도 6은 DM 정보의 일일수집모듈을 화면상에 구현한 프로그램을 캡처한 그림
- 도 7은 최신데이터수집모듈의 계정정보 수집 순서도

- 도 8은 최신데이터수집모듈의 FriendID 수집 순서도
- 도 9는 트윗 AIP 생성모듈의 순서도
- 도 10은 트윗 AIP 생성모듈의 화면 및 생성한 xml을 보기한 화면
- 도 11은 DM AIP 생성모듈의 순서도
- 도 12는 Favorite AIP 생성모듈의 순서도
- 도 13은 Follower AIP 생성모듈의 순서도
- 도 14는 Friend AIP 생성모듈의 순서도
- 도 15는 List AIP 생성모듈의 순서도
- 도 16은 보존 메타데이터 입력모듈의 순서도
- 도 17은 보존 메타데이터 입력 화면
- 도 18은 메타데이터 AIP 생성모듈의 순서도
- 도 19는 트위터 기록물의 장기보존 포맷
- 도 20은 트위터 기록물건 XSD
- 도 21은 트윗 문서의 XSD

발명을 실시하기 위한 구체적인 내용

- [0018] 이하 본 발명의 실시를 위한 구체적인 실시예를 도면을 참고하여 설명한다. 본 발명의 실시예는 하나의 발명을 설명하기 위한 것으로서 권리범위는 실시예에 한정되지 않고, 예시된 도면은 발명의 명확성을 위하여 핵심적인 내용만 확대 도시하고 부수적인 것은 생략하였으므로 도면에 한정하여 해석하여서는 아니 된다.
- [0019] 본 발명에서 트위터 데이터를 수집하기 위하여 트위터사에서 제공하는 API(Application Program Interface)인 REST(Representational State Transfer) API를 사용한다. REST API는 사용하기 쉽고 지원되는 라이브러리가 많으며, 데이터 수집시 사용되는 API는 Timeline, 트윗(Tweet), User, List, DM(Direct Message), Follow, Social Graph, Account, Favorite, Notification, Blocks, Spam Report, Help, OAuth, 저장 검색조건, 지역정보, 위치정보 관련 17가지이다.
- [0020] 트위터사에서 제공하는 API의 이용시의 문제점은 첫째, 아카이빙을 위한 데이터 수집시 일자별 정보수집이 불가능하고, 둘째, 트위터사의 정책에 의하여 트윗의 경우 최대 3,200개, DM의 경우 800개의 정보만이 가능하고 그 이전에 등록된 데이터는 수집이 불가능하며, 셋째, 접속자의 폭주로 인한 부하를 줄이기 위해 시간당 API 실행 횟수가 제한된다는 것이다.
- [0021] 본 발명에서는 위와 같은 문제점을 해결하기 위해 다음과 같은 방법을 사용한다. 첫째는 수집 데이터를 저장할 수 있는 RDBMS를 이용한다. 수집시에는 일자와 관계없이 최근 정보를 DB에 저장하고, 아카이빙용 데이터를 생성하기 위해 DB에서 일자별로 정보를 읽어 생성한다. 둘째는 트윗과 DM의 경우 최초로 일괄적으로 최대건수를 수집하는 모듈과 저장된 수집정보 DB에서 가장 최근의 트윗 ID를 읽은 후 그 이후에 등록된 정보만을 수집하는 모듈로 분리하여 구성한다. 셋째는 API호출 횟수를 카운트하여 횟수가 초과하는 경우에 스레드 지연(thread delay)방식을 이용하여 특정시간 이후 재호출하는 로직을 구성한다.
- [0022] 도 1은 본 발명에서의 데이터 수집 절차도를 보여준다. 트위터 데이터를 수집하기 위해서는 먼저 이용자의 ID와 비밀번호를 이용하여 인증정보를 받아야 한다. 인증방법으로는 BASIC, OAuth(Open Authorization), XAuth가 있으며, BASIC 인증은 API 호출횟수가 적고, XAuth는 DM을 수집하지 못한다는 단점이 있어 본 발명에서는 OAuth 인증방법을 사용한다.
- [0023] 구체적인 인증 절차는 다음과 같다. 트위터 수집모듈이 트위터 서버에게 임시로 사용되는 리퀘스트 토큰을 요청하면, 그에 대하여 상기 트위터 서버는 리퀘스트 토큰을 발급해 주게 된다. 이후 리퀘스트 토큰을 이용하여 인증을 수행하면 핀번호를 발급해 준다. 상기 트위터 수집모듈은 상기 핀번호를 받은 후 해당 핀번호와 리퀘스트 토큰 등의 정보를 이용하여 트위터 서버에게 액세스 토큰을 요청하게 되며, 상기 트위터 서버는 이러한 요청을 검증하여 해당 트위터 사용자의 리소스에 접근할 수 있는 액세스 토큰을 발급하게 되고, 상기 트위터 수집모듈

상기 액세스 토큰을 이용하여 사용자의 보호된 리소스에 접근할 수 있는 권한을 부여받게 된다.

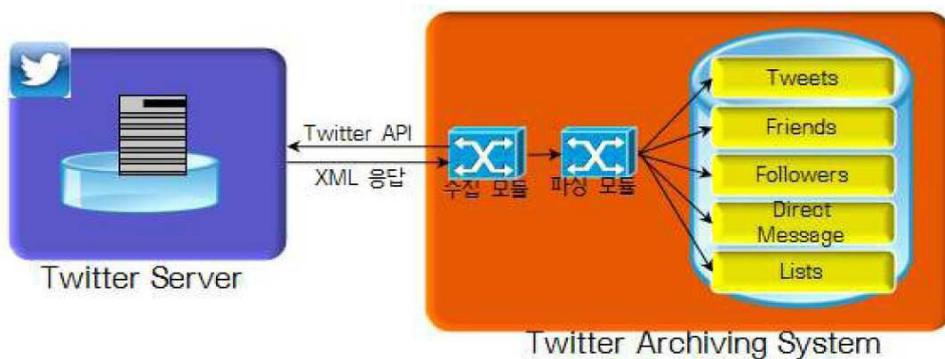
- [0024] 트위터 데이터 수집 절차는 트위터 API를 이용해 트위터 서버에 자료를 요청한다. 트위터 서버는 인증정보를 먼저 확인하고 API에 해당하는 결과값을 JSON(JavaScript Object Notation) 데이터 포맷으로 요청한 클라이언트인 수집모듈에 보낸다. 파싱(parsing) 모듈은 결과값을 파싱하여 수집정보 저장DB에 저장한다. 본 발명에서는 트위터사의 정책에 따라 API가 변경되거나 항목이 변화가 생길 것을 대비하여 유연성 있는 파서(parser)를 사용한다.
- [0025] 본 발명에서의 문서는 XML(extensible markup language) 기반의 데이터 파일로서 월드 와이드 웹페이지에 사용되며, 문서형식을 구조적으로 표현하고 검색하는 것이 용이하며, 재현이 가능하여 영구적 보존 파일로 사용한다.
- [0026] 본 발명에 따른 트위터 기록물 장기보존 시스템은 도 2와 같이 크게 수집부, 보관부, 재현부, 콘솔부 및 DB부로 구성된다.
- [0027] 상기 수집부는 트위터 기록물을 수집하는 도구로서 일괄수집모듈, 일일수집모듈 및 최신데이터수집모듈로 구성된다.
- [0028] 상기 일괄수집모듈은 행정기관에서 등록한 예전 자료를 최초로 수집하는 모듈로서 트위터사의 서버에 보관된 트윗 정보는 현재부터 3,200건의 예전 자료를 일괄적으로 수집하고, DM 정보는 현재부터 800건의 예전 자료를 일괄적으로 수집한다.
- [0029] 상기 일일수집모듈은 행정기관에서 등록한 최근의 트윗과 DM을 일일로 수집한다. 일일수집 대상이 되는 자료는 스케줄러(cron)에 의해 현재 보관된 트윗 및 DM보다 최근에 등록된 자료를 수집하여 DB부에 저장한다. 상기 일일수집모듈은 트윗과 DM이외에 팔로잉하고 있는 사용자 정보인 FriendID, 팔로워하고 있는 사용자 정보인 FollowerID 및 사용자정보를 수집 저장한다.
- [0030] 상기 최신데이터수집모듈은 아카이빙 직전의 최신 정보를 일회 수집하는 모듈로서, 행정기관의 계정정보, FriendID, FollowerID, 트윗을 리트윗하는 사용자 정보인 Retweet_of_Me, 관심글로 등록한 트윗 정보인 Favorite 및 그룹별 관리를 위해 등록한 List가 수집 대상이다.
- [0031] 도 3은 트윗 정보에 대한 일괄수집모듈의 알고리즘 순서도로서 인증절차를 거치고, 트윗 수집 함수를 호출하여 트위터 API를 호출하여 API 값에 해당하는 결과값을 얻어 데이터를 파싱하여 DB에 저장하고 저장된 정보가 없을 때(null)까지 반복 수행한다. DM 정보에 대해서도 같은 방법으로 일괄 수집한다.
- [0032] 도 4는 일괄수집모듈의 하나의 예로서 트윗 정보에 대한 일괄수집모듈을 화면상에 구현한 프로그램을 캡처한 것이다.
- [0033] 도 5는 트윗 정보에 대한 일일수집모듈의 알고리즘 순서도로서 일괄수집모듈의 알고리즘과 같고 다만 일별, 월별로 주기적으로 수집하는 것이 다르다. DM 정보에 대해서도 같은 방법으로 주기적으로 일일 수집한다.
- [0034] 도 6은 일일수집모듈의 하나의 예로서 DM 정보의 일일수집모듈을 화면상에 구현한 프로그램을 캡처한 것이다.
- [0035] 도 7은 최신데이터수집모듈의 계정정보 수집 순서도로서 아카이빙 직전에 행정기관의 계정정보를 계정 수집함수를 호출하여 수집한다.
- [0036] 도 8은 최신데이터수집모듈의 FriendID 수집 순서도로서 Friend 수집 함수를 호출하여 팔로잉하고 있는 사용자 정보를 수집한다. FollowerID, Retweet_of_Me, Favorite 및 List도 FriendID와 유사한 알고리즘으로 수집한다.
- [0037] 상기 보관부는 상기 수집부에서 수집한 정보를 구조적으로 표현하고 검색하기에 용이하며, 재현이 가능하여 영구적 보존 파일로 만들기 위하여 xml 생성하는 기능을 하는 것으로서, 트윗 AIP(Archive Information Package) 생성모듈, DM AIP 생성모듈, Following AIP 생성모듈, Follower AIP 생성모듈, Favorite 생성모듈, List 생성모듈, 보존 메타데이터 입력모듈 및 메타데이터 AIP 생성모듈로 구성된다.
- [0038] 도 9는 트윗 AIP 생성모듈의 순서도로서 수집한 트윗 정보와 리트윗 정보를 바탕으로 트윗 AIP를 생성하여 xml 데이터를 생성하고 보기를 할 수 있다.
- [0039] 도 10은 트윗 AIP 생성모듈의 화면을 캡처한 그림으로서 행정기관과 연도를 선택하여 생성버튼을 누르면 xml 데이터가 생성되고 보기버튼을 누르면 하단 화면과 같이 xml 데이터를 볼 수 있다.
- [0040] 도 11은 DM AIP 생성모듈의 순서도로서 수집한 DM 정보를 바탕으로 DM AIP를 생성하여 xml 데이터를 생성하고

보기를 할 수 있다.

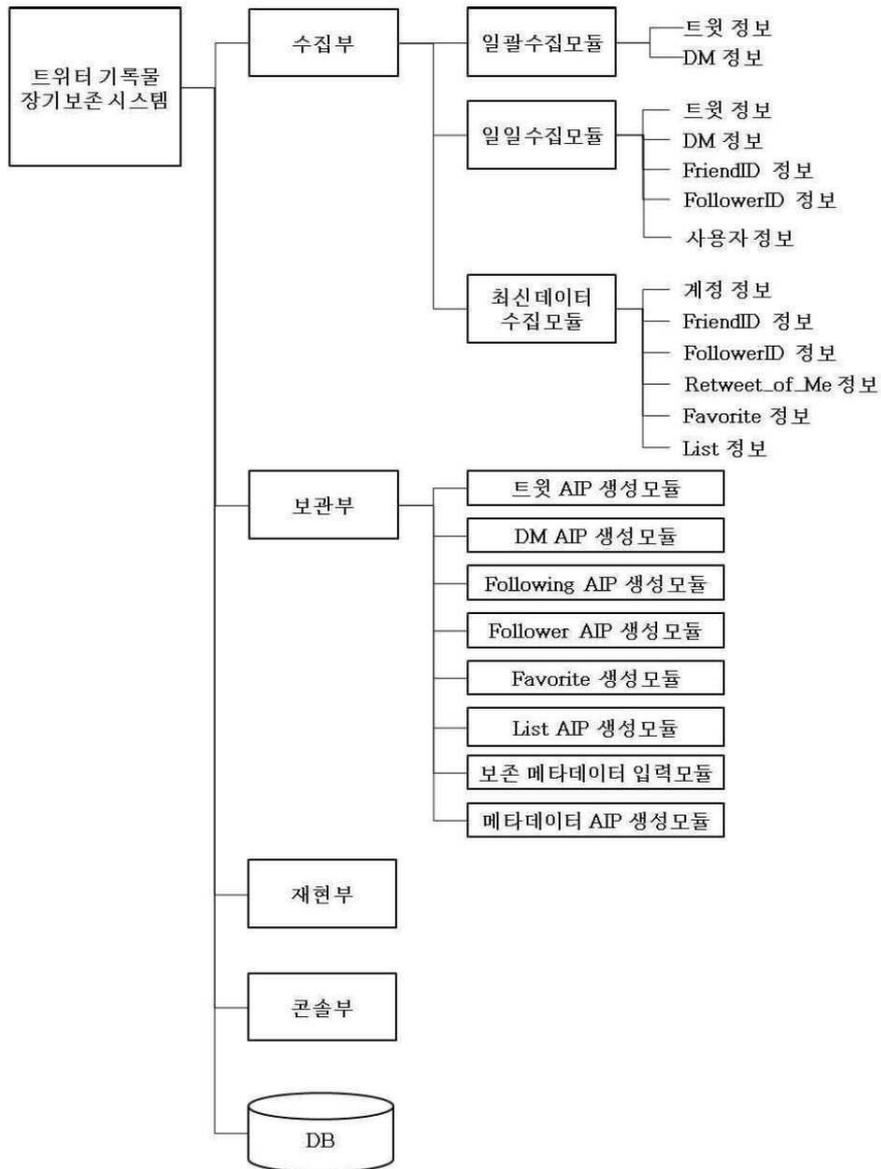
- [0041] 도 12는 Favorite AIP 생성모듈의 순서도로서 수집한 Favorite 정보를 바탕으로 Favorite AIP를 생성하여 xml 데이터를 생성하고 보기를 할 수 있다.
- [0042] 도 13은 Follower AIP 생성모듈의 순서도로서 수집한 Follower 정보를 바탕으로 Follower AIP를 생성하여 xml 데이터를 생성하고 보기를 할 수 있다.
- [0043] 도 14는 Friend AIP 생성모듈의 순서도로서 수집한 Friend 정보를 바탕으로 Friend AIP를 생성하여 xml 데이터를 생성하고 보기를 할 수 있다.
- [0044] 도 15는 List AIP 생성모듈의 순서도로서 수집한 List 정보를 바탕으로 List AIP를 생성하여 xml 데이터를 생성하고 보기를 할 수 있다.
- [0045] 도 16은 수집 및 보존 메타데이터 입력모듈의 순서도로서 메타데이터 AIP 생성을 위한 정보를 입력하는 모듈이다.
- [0046] 도 17은 보존 메타데이터 입력 화면으로서 행정기관과 연도를 선택하고 보존행위자, 키워드 및 기타정보를 입력하고 등록버튼을 누른다.
- [0047] 도 18은 메타데이터 AIP 생성모듈의 순서도로서 수집 및 입력한 정보를 바탕으로 메타데이터 AIP를 생성한다.
- [0048] 도 19는 트위터 기록물의 장기보존 포맷으로서 연단위의 기록물철과 기록물건으로 구성된다. 기록물건은 메타데이터 문서로서 트윗, DM, Friends, Followers, Lists, Favorites, 통계문서(status)로 구성된다. 트윗 문서는 트위터에 등록되는 트윗, 리트윗, 답글을 포함한다. 트윗 문서의 메타데이터는 ID, 생성일자, 생성자와 같은 트윗을 식별할 수 있는 정보와 StatusFlag, DocumentType과 같은 트윗의 구분 정보가 포함되며, 문서내용에는 트윗 내용, 리트윗 여부 등의 실제값이 포함된다. Friends와 Followers 문서는 트위터의 팔로잉, 팔로워 정보를 가지며, Favorites는 관심글 정보, Lists는 해당 행정기관의 리스트 정보, Status는 통계정보를 포함하고 있다.
- [0049] 도 20은 트위터 기록물건 XSD(Xml Schema Document)를 보여주는 것으로서, 트위터 문서 포맷(Twitter Document Format)은 Twitter Document Metadata와 Twitter Document Contents로 구성된다. Twitter Document Metadata는 메타데이터 정보를 담고 있으며, Twitter Document Contents는 Tweets, Friends, Followers, Favorites, Lists, DMs, Status 정보를 포함한다.
- [0050] 도 21은 트윗 문서의 XSD를 보여주는 것으로서 트윗 정보를 담고 있다. 트윗 문서는 Tweet_metadata와 Tweet_content로 구성되고, Tweet_metadata는 created_at, status_id, user_id, screen_name, status_flag, documenttype으로 되어 있고, Tweet_content는 text, source, truncated, in_reply_to_status_id, in_reply_to_user_id, in_reply_to_screen_name, retweet_count, retweeted 등으로 구성된다. DM, Friends, Followers, Lists, Favorites, 통계문서(status) 등도 유사한 구조로 구성된다.

도면

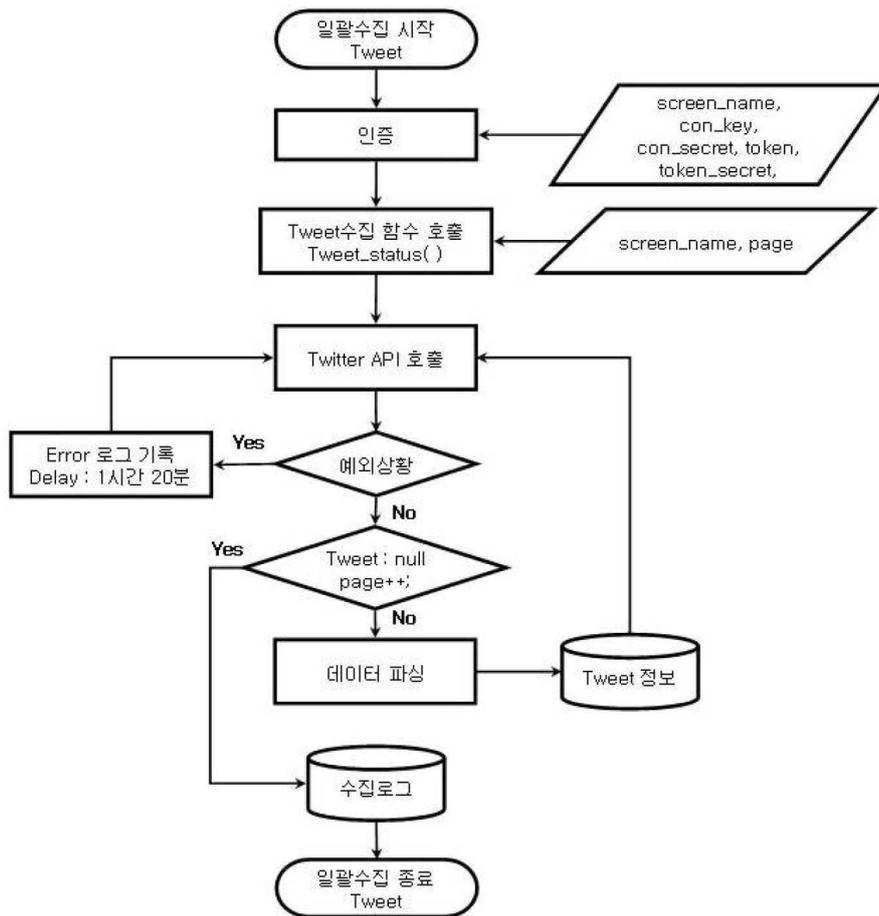
도면1



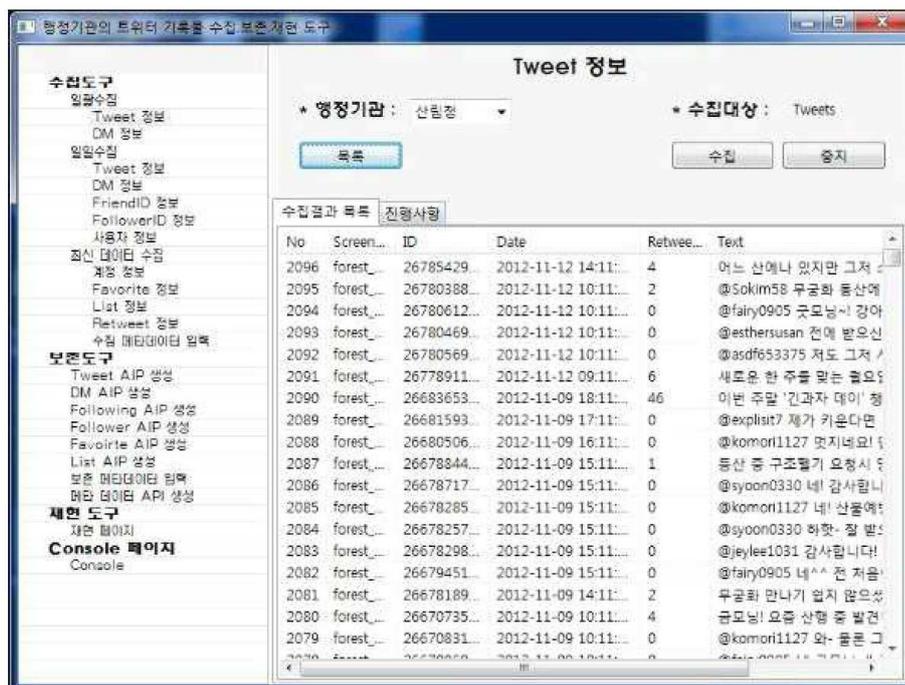
도면2



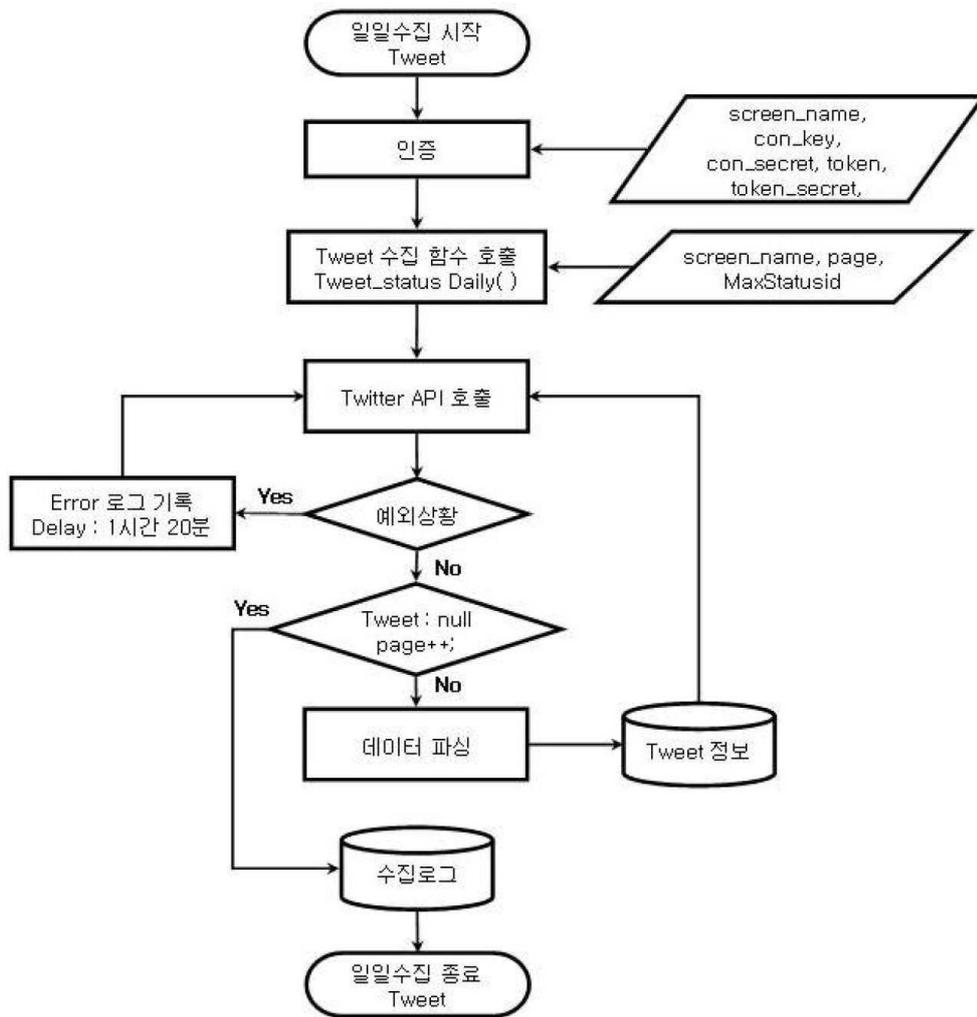
도면3



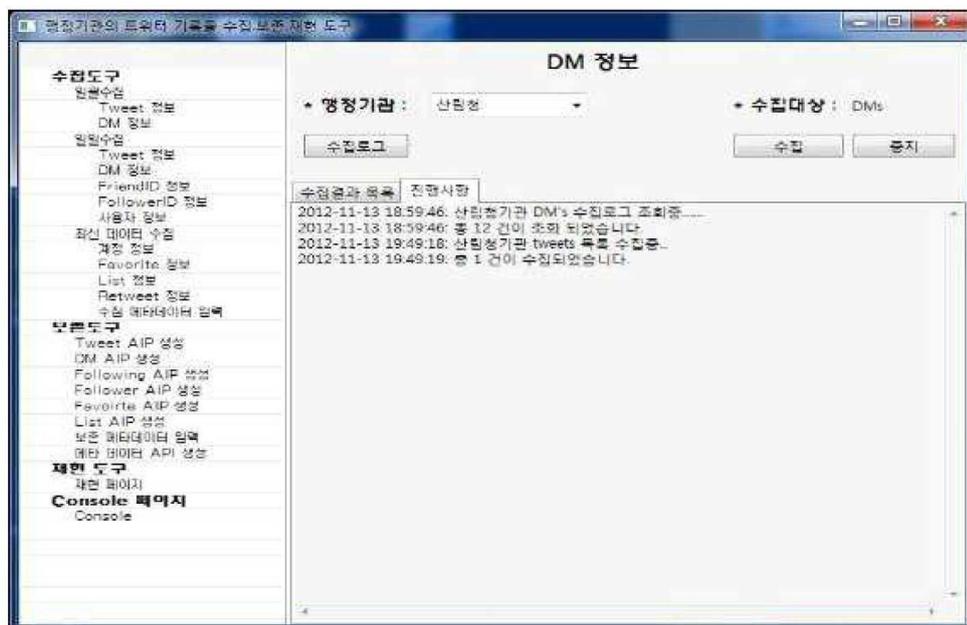
도면4



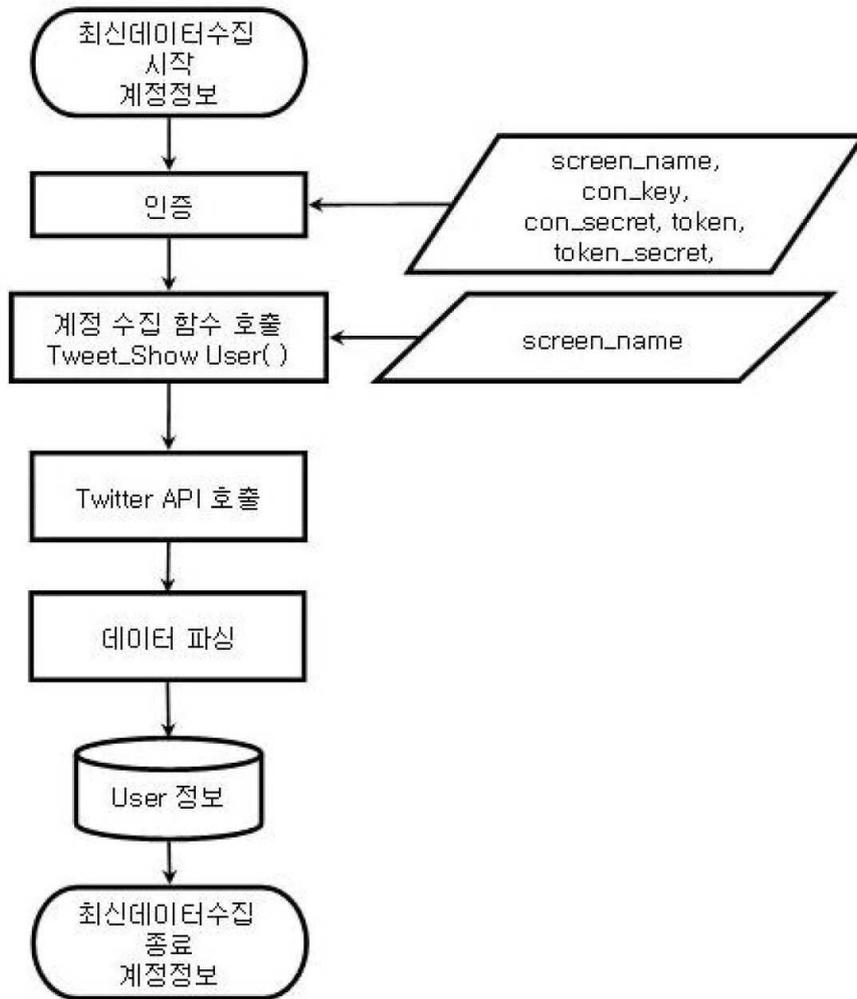
도면5



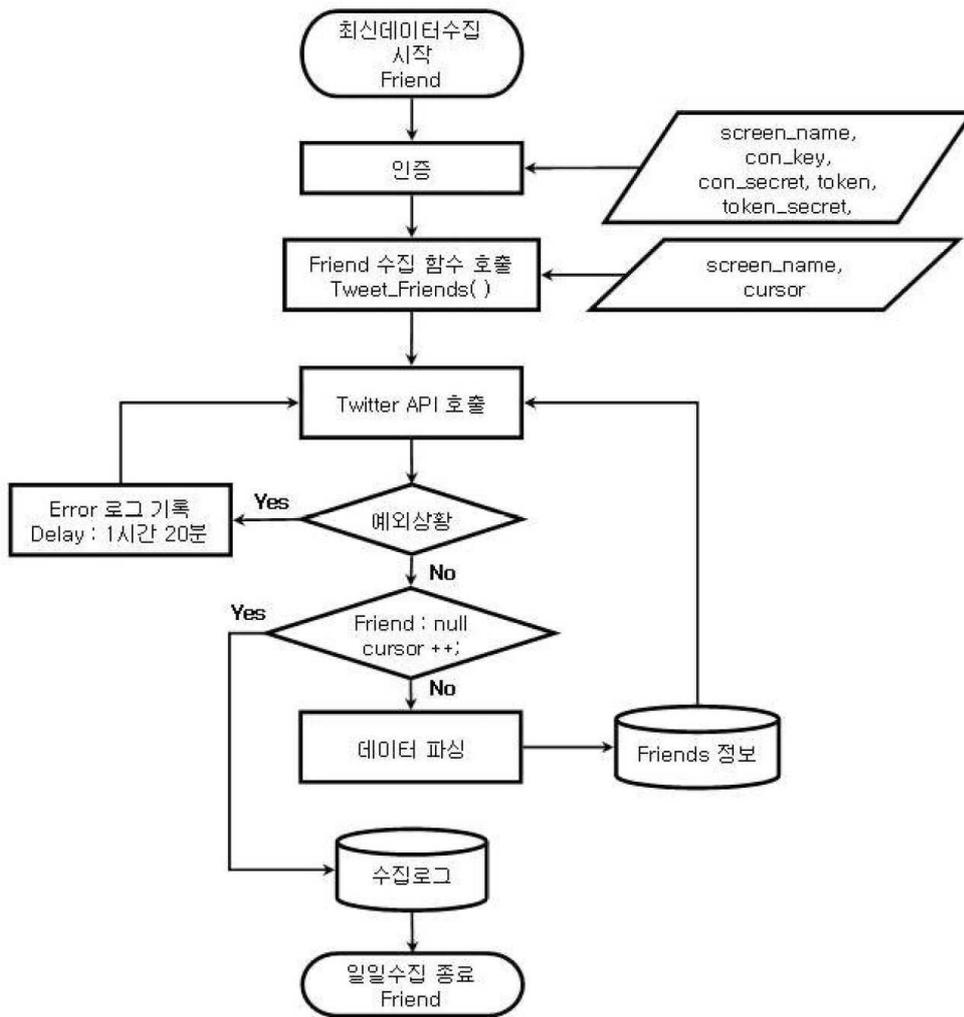
도면6



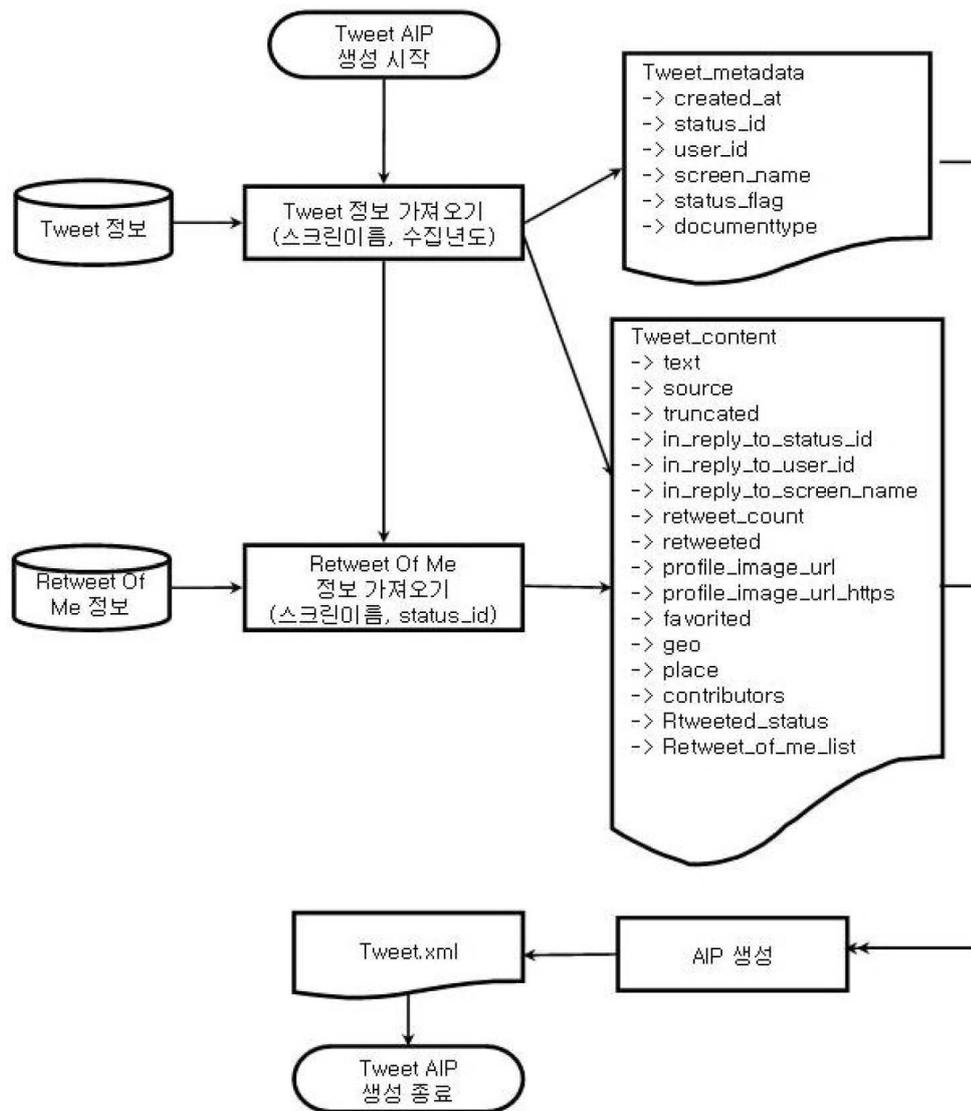
도면7



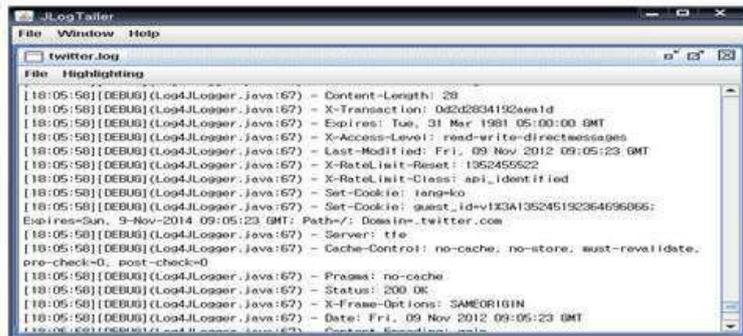
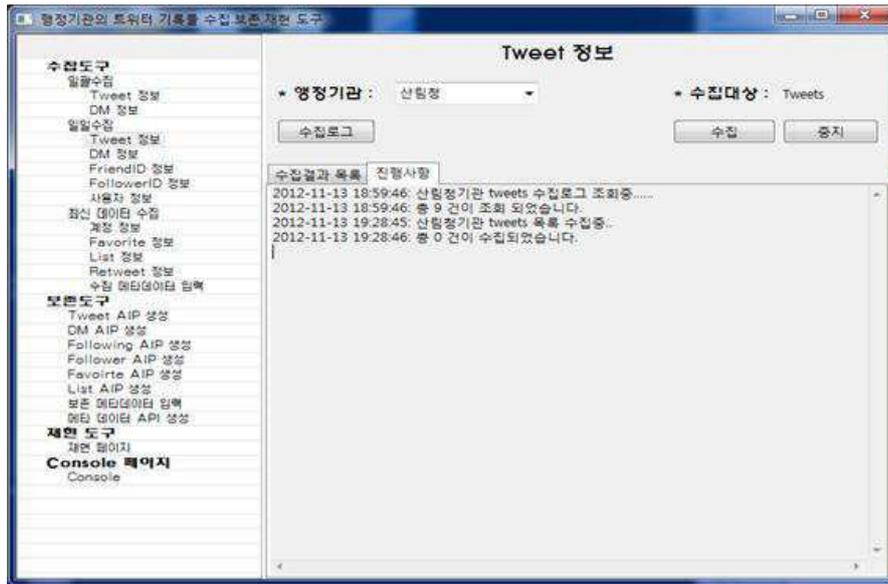
도면8



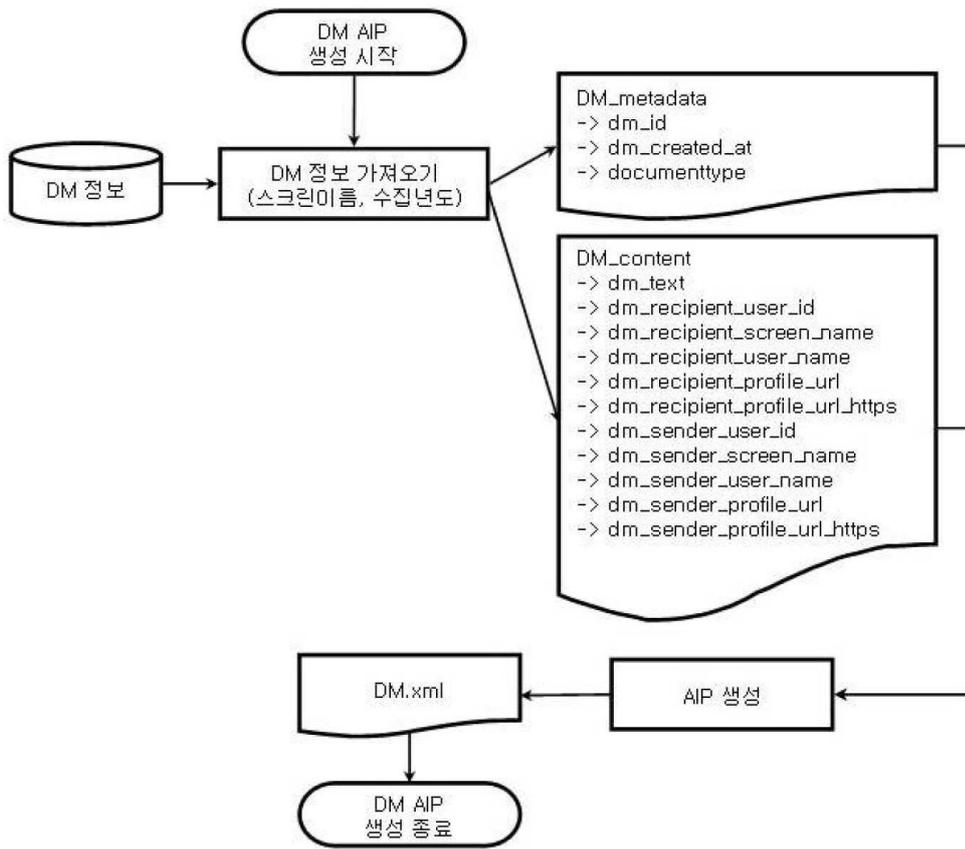
도면9



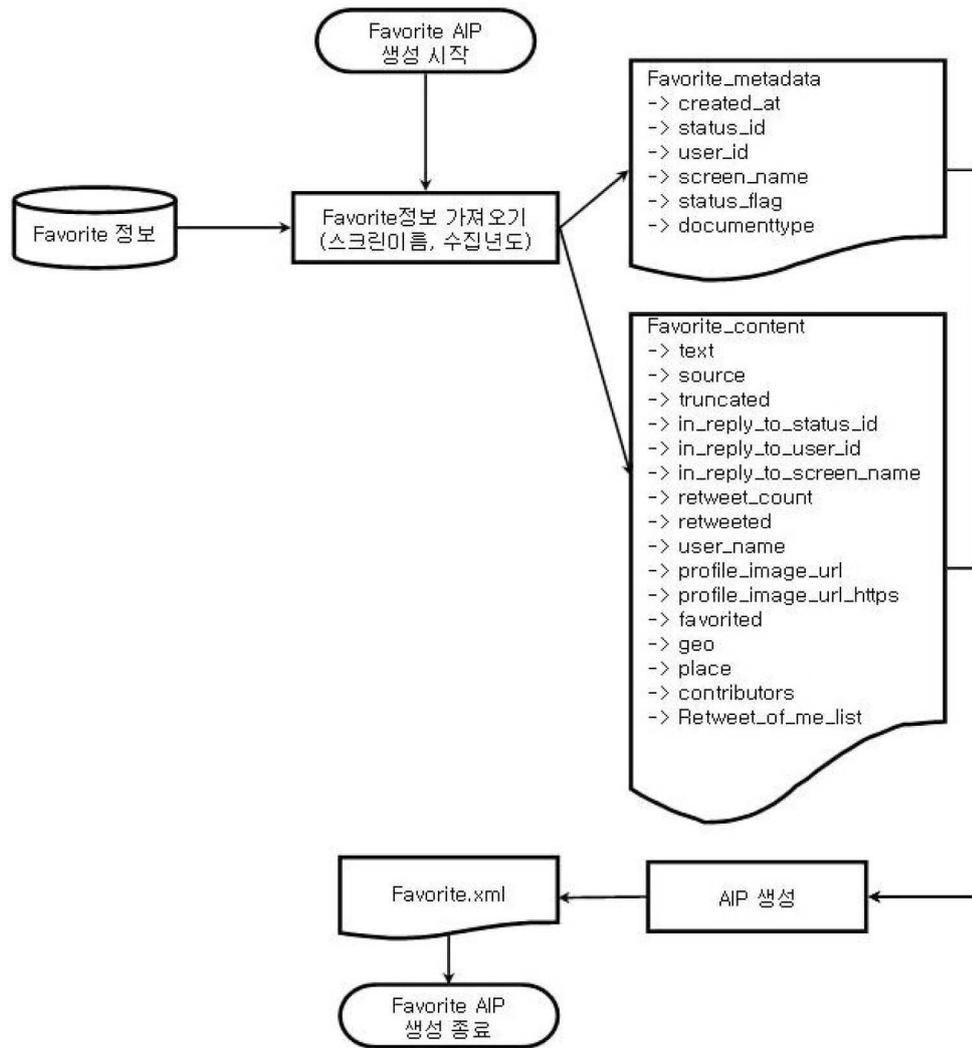
도면10



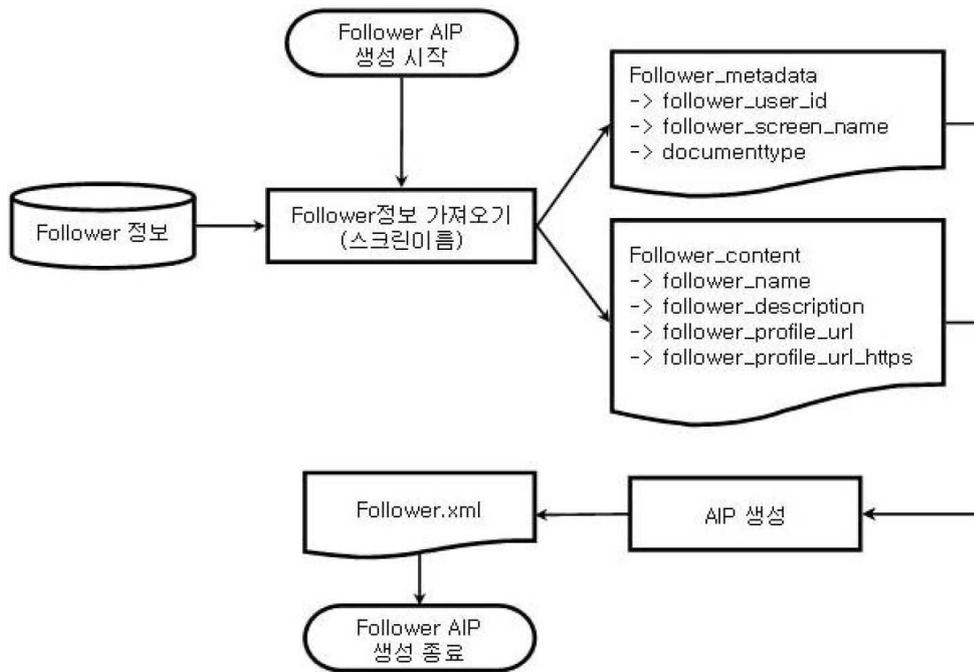
도면11



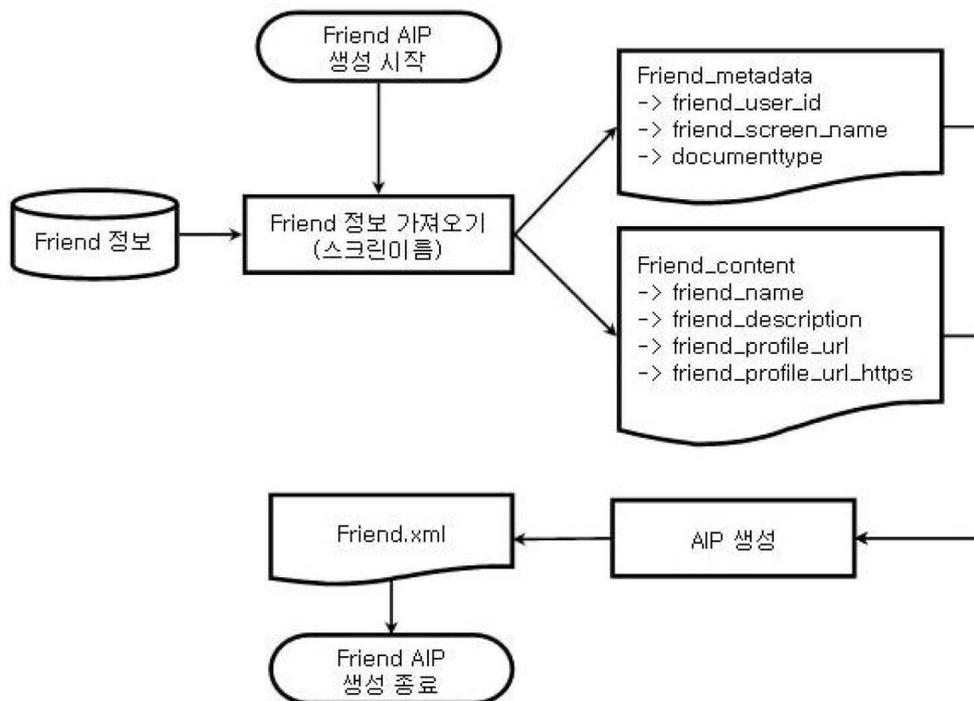
도면12



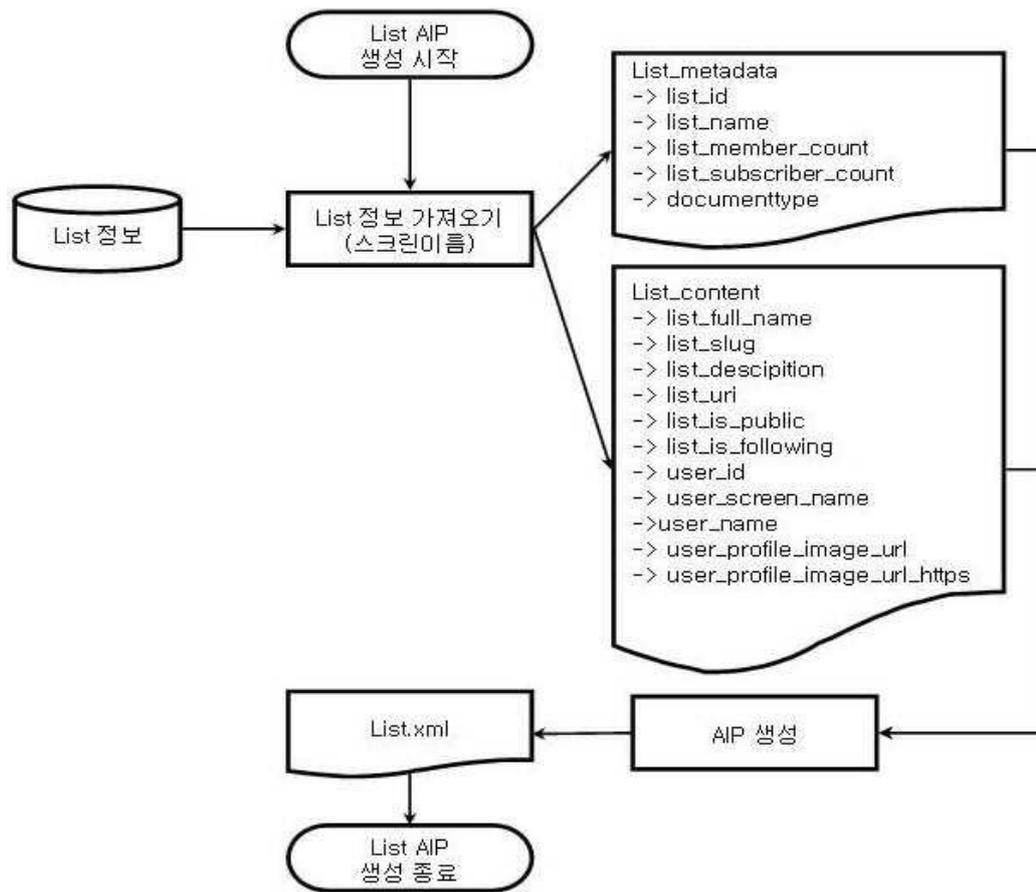
도면13



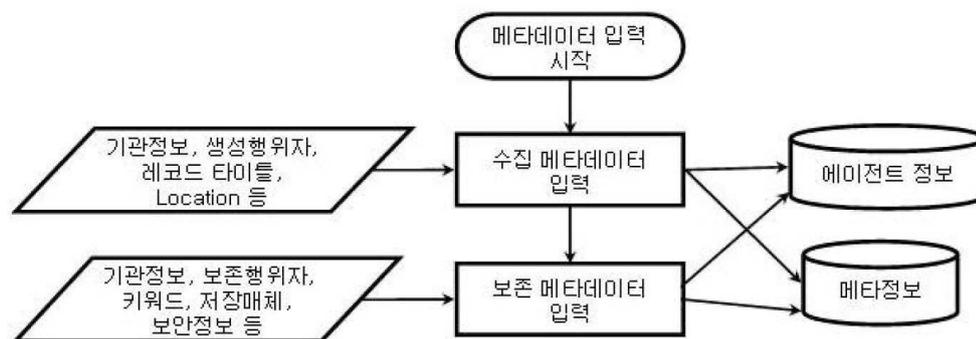
도면14



도면15



도면16



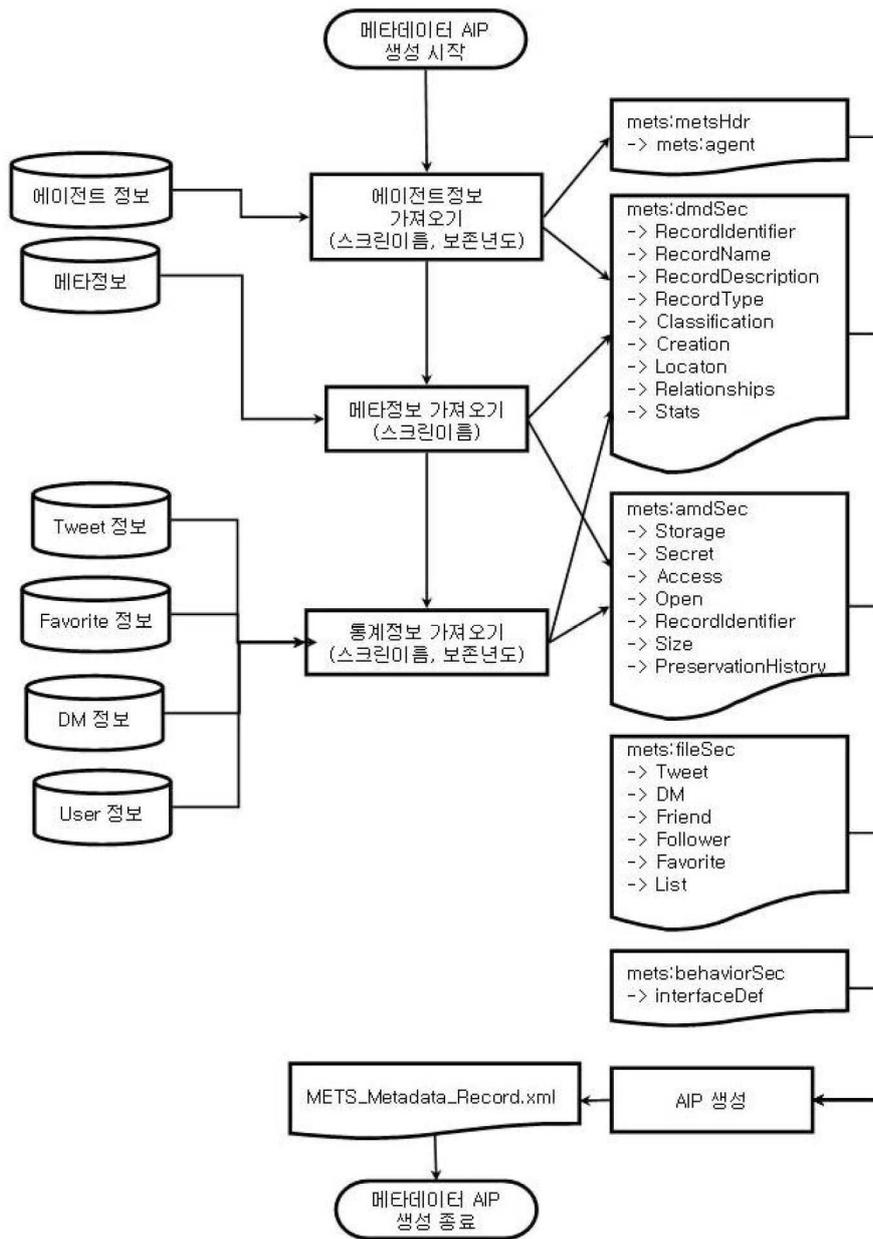
도면17

행정기관의 트위터 기록을 수집 보존 관련 도구

* 행정기관 : 산림청		* 보존년도 : 2012	
PreservationType	매체수출	PreservationDate	2012-11-13
Description	2012년 Tweet 데이터 보존		
ArchivorID	swhong	AgentType	보존행위자
Org_Name	에이티아이티	Org_Code	atit
PersonalName	홍삼옥	PersonalCode	090701
PartName	연구개발부	PartCode	pt01
PositionName		PositionCode	
Keywords			
RecordType	SNS 기록물	SNSType	트위터
StorageType	보존매체	StorageMedia	스토리지
ClassificationID		ClassificationName	
ClassificationLayer			
SecretType		SecretDueDate	
SecretReason			
AccessType	읽기	AccessTarget	전체열람
AccessMainName		AccessMainCode	
OpenRestrictionPart		OpenClassified	
ClosedReason		OpenSecheduleDate	
OpenReason			

등록

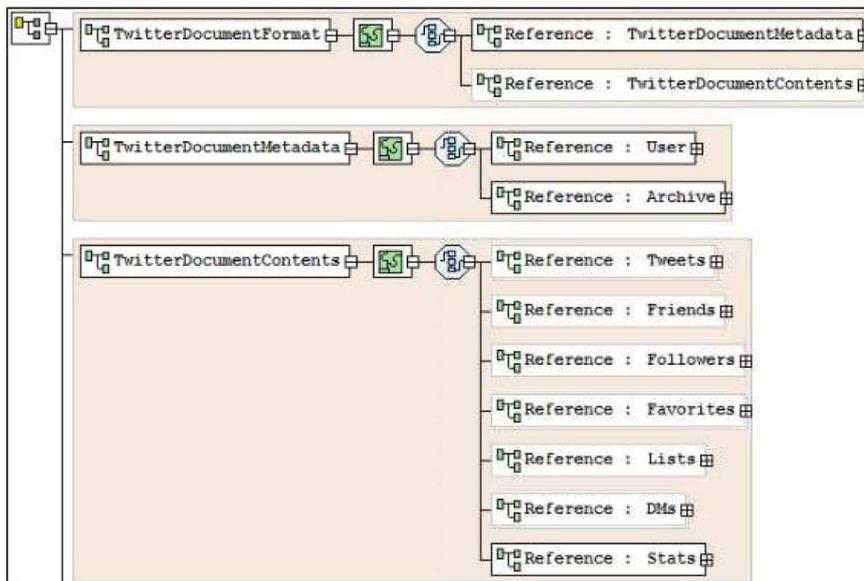
도면18



도면19



도면20



도면21

