



---

# 2018 국제인터넷보존컨소시엄(IIPC) 총회 및 컨퍼런스 참석 결과보고서

---

2018. 12.

# 목 차

I. 출장개요	1
II. 출장일정	2
III. IIPC개요	3
IV. 주요내용	4
V. 관련사진	19
VI. 종합의견	20
VII. 첨부자료	21

# I. 출장개요

---

## 1. 목 적

- 국제인터넷보존컨소시엄(International Internet Preservation Consortium: IIPC) 회원국('08.10 가입)으로서 회원 총회 및 컨퍼런스 참석
- 웹 아카이빙 관련 정보교환과 선진사례를 통해 오아시스(OASIS) 사업 개선 도모

2. 여행지 : 뉴질랜드(웰링턴, 뉴질랜드 국립도서관)

3. 참석자 : 김선미(디지털기획과 사서6급)

4. 기 간 : 2018. 11. 10.(토) ~ 11. 17.(토) [6박 8일]

## 5. 회의내용

- 회원 총회(1일)
  - IIPC 운영 및 회계 보고
  - 새로운 참가기관 소개
  - 워킹그룹 활동내용 보고 및 안건 토의
- 컨퍼런스(3일)
  - 각국의 웹 아카이빙 현황 발표
  - 웹 아카이빙 관련 주제발표 및 질의
- 공개 워크숍(1일)

## 6. 차기회의

- 기간/장소: 2019. 6. 5.~7. 크로아티아(자그레브)
- 주제 : Web Archiving community: Maturing practice together

## II. 출장일정

월일 (요일)	지역	시간	업무수행내용	비고
11. 10(토)	서울 뉴질랜드	17:05~08:15	<ul style="list-style-type: none"> <li>○ 인천 출발, 오클랜드 도착</li> <li>○ 오클랜드 출발, 웰링턴 도착</li> </ul>	
11. 11(일)		11:00~12:05		
11. 12(월)	뉴질랜드 (웰링턴)	09:00~18:00	<ul style="list-style-type: none"> <li>○ IIPC 총회 참가 등록</li> <li>○ 회원 총회 참석</li> <li>- 업무 및 회계 보고</li> <li>- 워킹그룹 보고</li> <li>○ 워킹그룹 회의 참석</li> </ul>	
11. 13(화)	뉴질랜드 (웰링턴)	08:30~20:00	<ul style="list-style-type: none"> <li>○ 컨퍼런스 참석</li> <li>- 개회식</li> <li>- 기조연설</li> <li>- 새로운 크롤링 접근</li> <li>- 접근 포털 및 API</li> <li>- 워크숍 및 Public Event</li> </ul>	
11. 14(수)	뉴질랜드 (웰링턴)	09:00~18:00	<ul style="list-style-type: none"> <li>○ 컨퍼런스 참석</li> <li>- 문화보존을 위해 진화하는 솔루션</li> <li>- 국내 협력, 디지털 보존</li> <li>- 웹 기록, 법적 고려사항</li> <li>- 워크숍</li> </ul>	
11. 15(목)	뉴질랜드 (웰링턴)	09:00~18:00	<ul style="list-style-type: none"> <li>○ 컨퍼런스 참석</li> <li>- 기조연설</li> <li>- 테마 수집 계획</li> <li>- 운영 워크플로</li> <li>- 웹아카이브 활용 등</li> <li>- 폐회</li> </ul>	
11. 16(금)	뉴질랜드 (웰링턴)	09:00~12:30	○ 공개 워크숍 참석	
11. 17(토)	뉴질랜드 서울	07:15~08:20 10:00~17:55	<ul style="list-style-type: none"> <li>○ 웰링턴 출발, 오클랜드 도착</li> <li>○ 오클랜드 출발, 인천 도착</li> </ul>	

### Ⅲ. IIPC(International Internet Preservation Consortium) 개요

1. 설립일자 : 2003년 6월(프랑스, 미국, 영국 국립도서관 등 12개 기관)

2. 설립목적

- 전 세계에서 생산되는 인터넷 정보자원을 수집·보존하고, 급변하는 정보환경에 대처방안 마련하여 장기적으로 접근을 가능하게 함
- 국제적인 아카이브 생성을 위한 공통의 도구(tool), 기술 및 표준을 개발하고 이용을 촉진하기 위함

3. 주요업무

- 웹 크롤러 개발과 데이터 저장형식과 같은 규격의 표준화
- 검색 및 자료 뷰잉시스템 개발 등 웹아카이브의 다양한 과제 연구
- 웹아카이브의 사례와 개발 성과 공유
- 수집, 검색, 보존 분야의 워킹그룹 운영 및 공동 연구 개발

4. 운영조직

- 의장 : Jefferson, Bailey(인터넷 아카이브)
- 부의장 : Sylvain Belanger(캐나다 국립도서관기록관)
- 운영위원회(15)

알렉산드리아 도서관, 영국 국립도서관, 콜롬비아대학교 도서관, 독일 국립도서관, 인터넷 아카이브, 캐나다 국립도서관기록관, 미의회 도서관, , 프랑스 국립도서관, 아이슬란드 국립대학교 도서관, 핀란드 국립도서관, 스페인 국립도서관, 네덜란드 국립도서관, 스탠포드대학교 도서관, 스위스 국립도서관, 노스텍사스 대학교

- 워킹그룹(3) : 콘텐츠 개발, 보존, 교육

5. 회원기관 : 56개관

구분	상세 내역	기관수
국가도서관	미국의회도서관, 국립중앙도서관 등	34
지역도서관	스페인 카탈로니아 등	2
대학 및 연구 도서관	스탠퍼드 대학 도서관 등	10
비영리재단 등	인터넷아카이브, 아카이브웹 등	4
국립문서보관소	영국국립문서보관소 등	2
기타 등	한조 아카이브, Mirrorweb 등	4
합 계		56

## IV. 주요 내용

---

### 1. 1일차 주요내용[회원 총회]

#### □ 개회식

- 환영사(Steve Knight, National Library of New Zealand)
- 개회사(Jefferson Bailey, Internet Archive)

#### □ 운영 보고

- 발표 : Jefferson Bailey(Internet Archive)
- 내용
  - IIPC 조직 상황 : 56개 회원(2018 신규 가입 회원 4개 기관)
  - IIPC 재정 상황 : 총자산 \$640,000(연수입 \$190,000, 연지출 \$640,000)
  - IIPC 커뮤니티 활동 : 콘텐츠 및 보존 워킹그룹의 지속적 활동, 교육 훈련 워킹그룹의 신설, 더 많은 협업 도구 개발 및 프로젝트(Slack, Meet-ups, etc) 수행
  - 새로운 프로젝트 : 회원 대상 의견조사 및 후속 활동, 포트폴리오 추가, 파트너십 & 대외관계(Outreach) 활동, IIPC 재량 펀딩 프로그램(DFP)

※ IIPC Discretionary Funding Programm(DFP) 재정 상황

- 총자산 \$640,000 + 연간 비축 금액 \$200,000
- 연간 운영 비용 \$100,000

- 향후계획 : 운영위원회(Steering Committee) 선거, 회원 및 커뮤니티 이익에 대한 집중, 성장 중심, 전략적 투자

※회원 활동에 대한 투자

- 재량 활동에 대해 \$550,000 제안
- 중요하고 다양한 투자를 위한 2-3년 투자 계획 제안
- 생산적이고 지속가능하며 전략적인 방법으로 회원의 업무를 발전시키고자하는 운영위원회의 의도를 회원에게 전달

- DFP와 IIPC 재정
  - 총자산 \$640,000 + 연간 비축 자산 \$200,000 + 연간 비용 \$100,000
  - 긴급 운용자금으로 연간 \$100,000 투자
- 회원활동을 위한 전략적 투자
  - 전용활동과 임시활동에 대해 \$550,000 투자 제안
  - 3년 투자 계획 : 1년차 \$250,000 ; 2년차 \$175,000 ; 3년차 \$100,000
    - ✓ Seed 교부금 : \$10,000 이상
    - ✓ 개발 교부금 : \$25,000 이상
    - ✓ 프로그램 교부금 : \$50,000 이상
- 신청기준 : IIPC 전략적 목표에 연결된 2개 회원이 포함되어야 하며, 다년간이 될 수도 있음

## □ 워킹그룹 활동보고 및 회의

- 콘텐츠 개발 워킹그룹(의장: Nicola Bingham, Alex Thurman)
  - 내용 : 컬렉션 업데이트 현황 보고(2018 동계 올림픽(1,413건), 1차 세계대전 기념 행사(2,539건), 국제 협력 조직(147건), 세계 주요 뉴스(1,453건)), 2019년 컬렉션 주제 제안, 예산 기획 등
- 보존 워킹그룹(의장: Tobias Steinke, Grace Thomas)
  - 내용 : OAIS 상호 운용 가능한 데이터 아키텍처 및 프로토콜 계획 보고 등
- 교육 훈련 워킹그룹(의장: Tom Cramer, Allie Grotke, Maria Praetzellis)
  - 내용 : 회원 대상 교육훈련에 대한 설문조사 결과 보고
    - 224명 응답(국립도서관 22%, 연구기관 40%, 대학 등 기타기관 38%)
    - 대부분의 사람들이 소프트웨어 개발을 제외한 모든 일을 담당
    - 응답자의 62%가 온라인 자원을 통해 교육 및 훈련 실시
    - 훈련 우선순위 : 웹아카이빙/디지털 보존 표준 및 기술
    - 훈련 채널 : 면대면 교육 > Webinar > 온라인매체
    - 미래 활동에 우선순위 : 장서개발(3.7) > 다른 회원과 함께 프로젝트 참여(3.4) > 컨퍼런스(2.9) > 커뮤니티 구축(2.8) > 도구 및 표준 개발(1.8)

- 기타의견 : Heritrix, Open Wayback 같은 개발 프로젝트를 위한 기금조성, 이벤트 커뮤니티 구축, 정보 공유 및 의사소통을 위한 플랫폼 개발
- 다양한 수준의 웹 아카이빙 훈련에 관심이 많은 것으로 조사되어 3수준의 커리큘럼 개발 예정 : 초급과정(웹 아카이빙 개요) ; 중급과정(도구) ; 고급과정(전문가 토폭)
- 설문조사 결과를 토대로 관심 주제별로 그룹을 나누어 정보 교류 (큐레이션, 콘텐츠 수집, 접근가능성, 컬렉션 분석, 다른 콘텐츠 수집 등)

## 2. 2일차 주요내용[컨퍼런스]

### □ Heritrix를 통한 고품질의 지속적이고 확장 가능한 웹 크롤링

- 발표 : Andrew Jackson(The British Library)
- 내용
  - 2013년 법적 납본 이후 크롤링 규모와 복잡성은 크게 증가하였으나 인적, 물적 지원은 증가하지 않아 한정된 자원으로 효율적으로 크롤링하는 방법에 대해 고민 시작. Heritrix3 설치가 많은 수의 별도 크롤링 작업을 실행할 때 작동이 번거롭고, 크롤링 프로세스 및 크롤링 상태 관리 방식이 작업 진행 상황을 파악하는데 어려움이 있어 이러한 문제점을 해결하기 위한 전략을 수립하고 실험에 착수
  - 실험 목표 : 높은 수준의 자동화, 높은 수준의 투명도, 자바스크립트가 많은 웹페이지 캡처, 안정적인 도메인 크롤링, 지속적인 크롤링
  - 문제점 해결을 위한 전략 : Heritrix에서 독점적으로 관리하는 크롤링 상태의 양을 줄이고 단일 크롤링에서 수백 개의 사이트를 독립적으로 크롤링할 수 있는 지속적인 크롤링 모델로 전환하여 결과적으로 크롤링 프로세스의 품질과 견고성이 크게 향상되었으며 Heritrix3 자체는 최소한으로 변경됨.



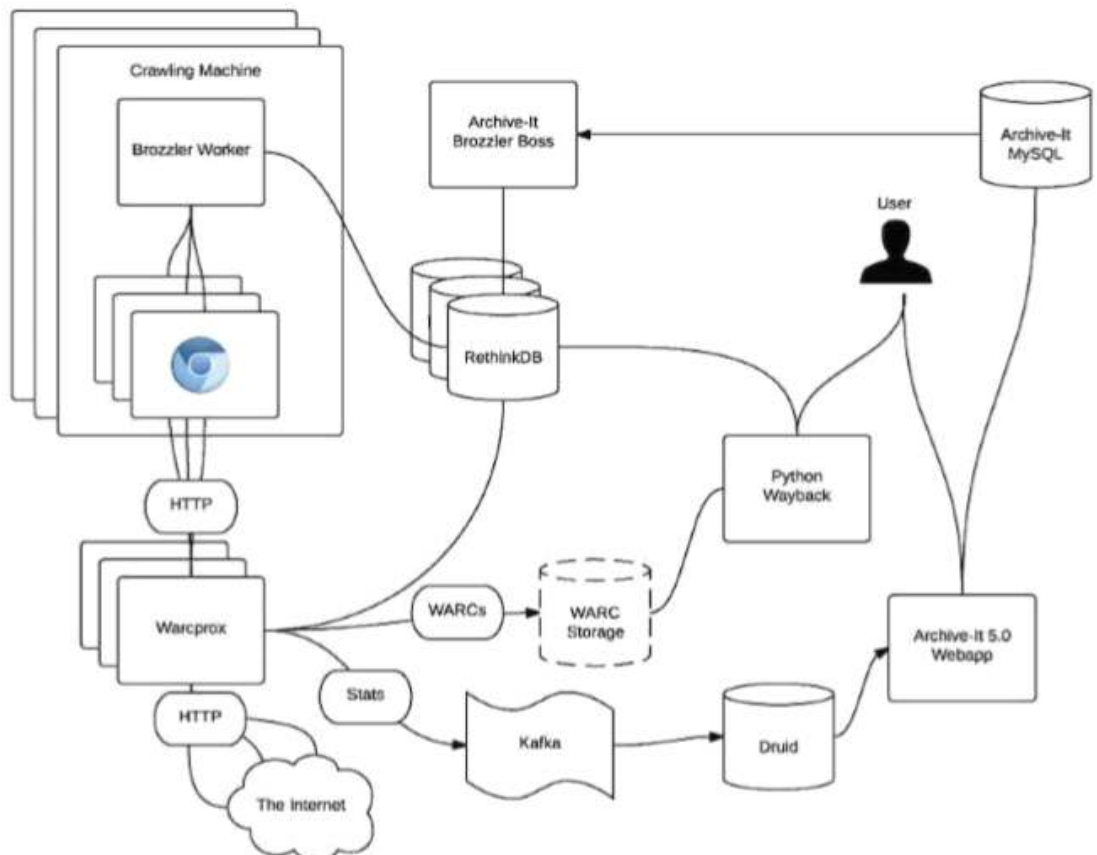
## □ 스트리밍 미디어가 포함된 웹사이트 아카이빙

○ 발표 : Howard Besser(New York University)

○ 내용

- 기존 웹 크롤링 소프트웨어(Heritrix와 Archive-It)는 스트리밍 미디어 수집에 취약
- 뉴욕대학도서관과 Internet Archive는 앤드류 멜론 재단으로부터 자금을 지원받아 협업을 통해 스트리밍 오디오 및 비디오 수집을 위한 새로운 크롤러 “Brozzler” 를 개발하여 현대 젊은 작곡가들의 웹 사이트 아카이빙에 테스트 적용
- 현재 웹크롤러의 결함을 조사하고 기존 웹 아카이빙 도구 및 서비스를 확장하여 새로운 크롤러로 오디오 및 비디오 스트림을 수집

<Brozzler System Architecture>



- 성공적인 프로젝트 수행을 위해 작곡가들과의 관계 구축 체계화(사이트 기증에 대한 동의서 및 계약서 작성 등), 프로젝트 결과에 대한 연구자 만족도 평가 실시
- 프로젝트 수행 결과 여전히 동영상 캡처 실패 등 일부 문제를 포함하고 있으나 스트리밍 미디어 캡처를 위한 도구로 Heritrix의 대안이 될 수 있음을 확인. 프로젝트 결과는 뉴욕대학교 도서관과 Archive-It 웹 사이트에서 확인 가능.

## □ 웹 아카이브 크롤링으로부터 이벤트 컬렉션 만들기

- 발표 : Martin Klein(Research Library)

Lyudmila Balakireva(Los Alamos National Laboratory)

Herbert Van De Sompel(Data Archiving and Networked Services The Netherlands)

- 내용

- 이벤트 중심의 웹 컬렉션은 주로 Archive-It과 같은 서비스로 라이브 웹을 크롤링하여 구축하고 있으나 Seed URI가 수작업으로 수집되는 경우가 많으며 관련성 여부가 사람의 평가에 의존한다는 단점이 있음
- 발표자들은 Memento 인프라를 통한 웹 아카이브 사용, Wikipedia 페이지에서 참조 사용, 콘텐츠 및 시간적 관련성 평가에 초점을 맞춘 크롤링 사용을 통해 웹 아카이브를 집중적으로 크롤링하여 이벤트 컬렉션을 구축하는 프로젝트 실시
- 결론 : 웹 아카이브는 웹 자원의 이벤트 컬렉션을 구축하는데 유용한 자원이 됨. 최근 이벤트에 대한 컬렉션은 아카이브웹보다 라이브 웹에서 더 많은 이점이 있으나 과거 이벤트에 대한 컬렉션은 라이브 웹보다 아카이브웹에서 더 많은 이점을 발견할 수 있음. 다양한 웹 아카이브를 활용하는 것이 컬렉션에 도움이 되며 집중 크롤링이 수동 수집보다 우수한 잠재력을 가짐.

## □ 뉴질랜드 국립도서관의 소셜미디어 수집

- 발표 : Gillian Lee(National Library of New Zealand)
- 내용
  - 소셜미디어 수집 동기 : 2011년부터 2017년까지 선거 캠페인에서 운영한 웹사이트와 소셜미디어를 비교한 결과 최근 웹사이트보다 소셜미디어를 활용하는 추세 확인
  - 2007년부터 웹크레이터 도구를 사용하여 웹사이트를 수집해 왔으나 소셜미디어 수집에 적합하지 않다고 판단, 많은 연구자들이 Twitter API를 사용하여 콘텐츠를 캡처하거나 연구 내용을 분석하는 것을 고려하여 Twitter의 API를 사용하여 트위터를 수집하는 두 번의 파일럿 프로젝트 진행
    - 2016년 카이 코우 라 지진 트위터 크롤링 : 해시 태그와 검색어가 포함된 중요한 사건과 관련하여 Twitter 데이터 캡처
    - 2017년 총선거 트위터 크롤링 : 모든 후보자와 정당의 트위터 계정을 크롤링. 해시 태그와 검색어를 포함(295개 계정 수집)

## □ 프랑스의 소셜미디어 수집

- 발표 : Géraldine Camille(Bibliothèque Nationale de France, BnF)  
Jérôme Thièvre(Institut National de l' Audiovisuel, INA)
- 발표내용
  - BnF는 선거 또는 올림픽 같은 특정 프로젝트를 위해 2011년에 트위터 크롤링 시작. 2018년 초부터 BnF 수집 정책에 따라 선택된 계정에 대해 일 년 내내 매일 두 차례씩 크롤링 시작.(2012년 1,033계정, 2017년 3,689계정 수집)
  - BnF의 목표는 최대한 많은 수의 트윗과 다른 계정 또는 해시태그와의 링크를 획득하여 라이브 웹에 게시할 때 제공하는 것이므로 API 대신

Heritrix3를 사용하여 트윗을 수집하였으며 수집 결과 동영상과 외부 링크들이 누락되는 한계를 보임

- INA는 2014년 공개된 트위터 API를 기반으로 트위터 크롤링 시작  
(이 공개 API는 실시간 트윗 검색에 제한이 있을 뿐만 아니라 기록된 데이터 검색도 더 어려워지는 제한사항은 있음.)
- INA는 4년 동안 약 1천 만 개의 트윗(tweets) 수집, 현재 12,000개의 계정과 560개의 해시태그를 지속적으로 크롤링하고 비디오와 이미지도 크롤링함
- INA의 다음 목표 중 하나는 TV아카이브와 함께 트위터 아카이브가 이용 가능하도록 만드는 것이므로 INA는 시청각 프로그램과 관련된 트윗 데이터를 매쉬업 할 소셜TV 사용자 인터페이스를 개발 중임

### 3. 3일차 주요내용(컨퍼런스)

#### □ End of Term(EOT) 아카이브 : 미국 정부 웹 공동 보존

- 발표 : Abbie Grotke(Library of Congress)

Mark Phillips(University of North Texas Libraries)

- 내용

- (개요) EOT 프로젝트는 미국 정부 전환 시기 웹사이트를 아카이빙하기 위해 미의회도서관, 인터넷 아카이브 등 6개 기관이 공동으로 추진해온 프로젝트로 2008년, 2012년에 이어 2016~2017년에 세 번째 프로젝트 진행
- (목표) 공개된 미국 정부 웹사이트의 보존을 위한 협업, 대통령 직무 기간 중 연방 정부 기관의 웹상에서의 존재를 문서화, 파트너 기관의 기존 연구 컬렉션 강화, 보존 필요성에 대한 인식 제고, 연구원 및 주제 전문가와의 교류

- (내용) EOT 2016의 결과로 9,000개 이상의 소셜미디어 계정(페이스북 44%, 트위터 37%, 유튜브 10%), 190,000개 이상의 도메인, 서브도메인, gov/nongov 사이트 수집

※ EOT 컬렉션 결과

2008년	2012년	2016년
16TB 160million URLs/files	35TB 194million URLs/files	250TB 350million URLs/files

□ 공개 기록 보존과 잊혀질 권리

- 발표 : Nicola Bingham(The British Library)
- 내용
  - UK Web의 사명은 포괄적이고 가능한 한 많이 접근할 수 있는 웹 컬렉션을 만드는 것이며 이는 책임감 있게 합법적, 윤리적으로 성취하여야 함
  - 프라이버시와 민감한 개인 정보 온라인 노출 위험에 대한 대중의 관심 증가하고 있으며, EU의 GDPR<sup>1)</sup>과 새로운 영국 전용 데이터보호법이 웹 아카이빙에 영향을 미침
  - 공공기관은 공공의 이익을 위해 수행하는 업무에 대해 면책 가능성이 있으나, 데이터 주체가 우선순위를 갖고 있어 중대한 피해나 고통을 주장하며 정보 삭제를 요청할 수 있음. Take down은 삭제보다는 접근 일시 중단으로 해석함.
  - Take down을 위한 제안 : 웹아카이빙 운영위원회 구성, Take down 정책 검토를 위한 명확한 지배 구조 구현, 현장 직원을 위한 다양한 시나리오 및 응답을 제시하는 가이드라인 명문화, 자료의 빠른 Take down을 위한 기술적인 과정 검토

1) 유럽 일반 개인정보보호법(General Data Protection Regulation). 2018년 5월 25일부터 시행된 EU의 개인정보보호 법령이며, 동 법령 위반시 과징금 등 행정처분 부과(출처: 개인정보보호 종합포털, www.privacy.go.kr)

- 2013년 이후 40건 정도의 Take down요청 발생(요청 사유: 데이터 보호 위반, 잠재적으로 민감한 개인정보, 명예훼손죄, 저작권 이슈, 부정확한 데이터, 웹 아카이빙에 대한 오해)
- 명예훼손 자료는 Take down(서비스 중지), 부정확한 자료나 분쟁 사실은 일반적으로 서비스 유지(단, 건강, 안전 및 보안에 관한 사항은 제외)
- 오픈 액세스 라이선스 확보(저작권 침해 주장으로부터 보호하기 위해 송장 작성)
- 다음과 같은 개인정보의 경우 게시 중단할 가능성이 낮음  
개인에게 당혹스러운 상황일 수는 있으나 손해가 되지는 않는 자료, 장기간 널리 이용될 수 있는 자료, 공적으로 책임이 있는 인물에 대한 것, 아카이브된 복사본이 유일한 복사본일 경우
- 특별한 고려가 필요한 경우 : 어린이나 취약계층에 대한 자료, 프라이버시가 보안문제가 될 때, 주제지식이나 동의 없이 이용 가능한 자료, 심각한 우울증이나 정신적 고통을 겪을 때 포스트 된 자료
- Take down 가이드라인이 현장 직원에게 도움이 되지만 표준화된 응답이 항상 최선은 아니므로, 소규모 관리 그룹이 어려운 요청을 다루고 정책을 알리기 위해 Take down 요청을 데이터베이스화 하고, 정기적으로 검토할 것을 제안

## □ [튜토리얼] 웹 큐레이터 도구 리런치

- 워크숍 코디네이터 : Ben O'Brien, Hanna Koppelaar
- 내용
  - 뉴질랜드 국립 도서관 (NLNZ)과 네덜란드 국립 도서관 (KB-NL)의 협업을 통해 2018년 1월부터 추가 된 Web Curator Tool(WCT) 새로운 기능 소개
  - WCT를 현대 웹 보관 요구사항에 부합할 수 있는 플랫폼으로 가져 오는 것이 목표

□ **[워크숍] WARC 파일 형식 : 다음 단계 준비**

- 워크숍 코디네이터 : Sara Aubry
- 내용 : 2020년 예정인 WARC 표준 개정에 관한 정보 및 의견 공유

※ WARC 파일 형식 ISO 제·개정 추진경과

- 2009년 5월 ISO 국제 표준(28500:2009 (WARC 1.0))으로 출시
- 2017년 8월 ISO 28500:2017(WARC 1.1)로 개정
- 2020년 개정을 위한 ISO 투표 예정

□ **[튜토리얼] 개인 및 기관을 위한 인간 규모 웹 수집(Webrecorder tutorial)**

- 튜토리얼 코디네이터 : Anna Perricci
- 내용 : Webrecorder.io로 웹 아카이브 구축, 관리 및 공유하는 방법, Webrecorder Player(WARC 파일접근을 위한 데스크톱 응용 프로그램)를 사용하여 오프라인에서 웹 아카이브와 상호 작용하는 방법에 대한 실무 지식 제공

□ **[튜토리얼] 협업하는 웹 아카이브 수집 프로젝트 관리를 위한 Cobweb 사용**

- 워크숍 코디네이터 : Kathryn Stine, Stephen Abrams, Peter Broadwell
- 내용 : Cobweb 계정 설정, 수집 프로젝트 수립 및 갱신, 기술적인 메타데이터 할당, Cobweb 레지스트리 내에서의 고급 검색 기능, Cobweb의 다중 참여자 커뮤니케이션 지원 방법, Cobweb 플랫폼의 지속적인 유지 관리 및 향상을 위한 로드맵 등 설명

## 4. 4일차 주요내용(컨퍼런스)

### □ 네덜란드 정부기관 웹 사이트 아카이빙을 위한 가이드라인

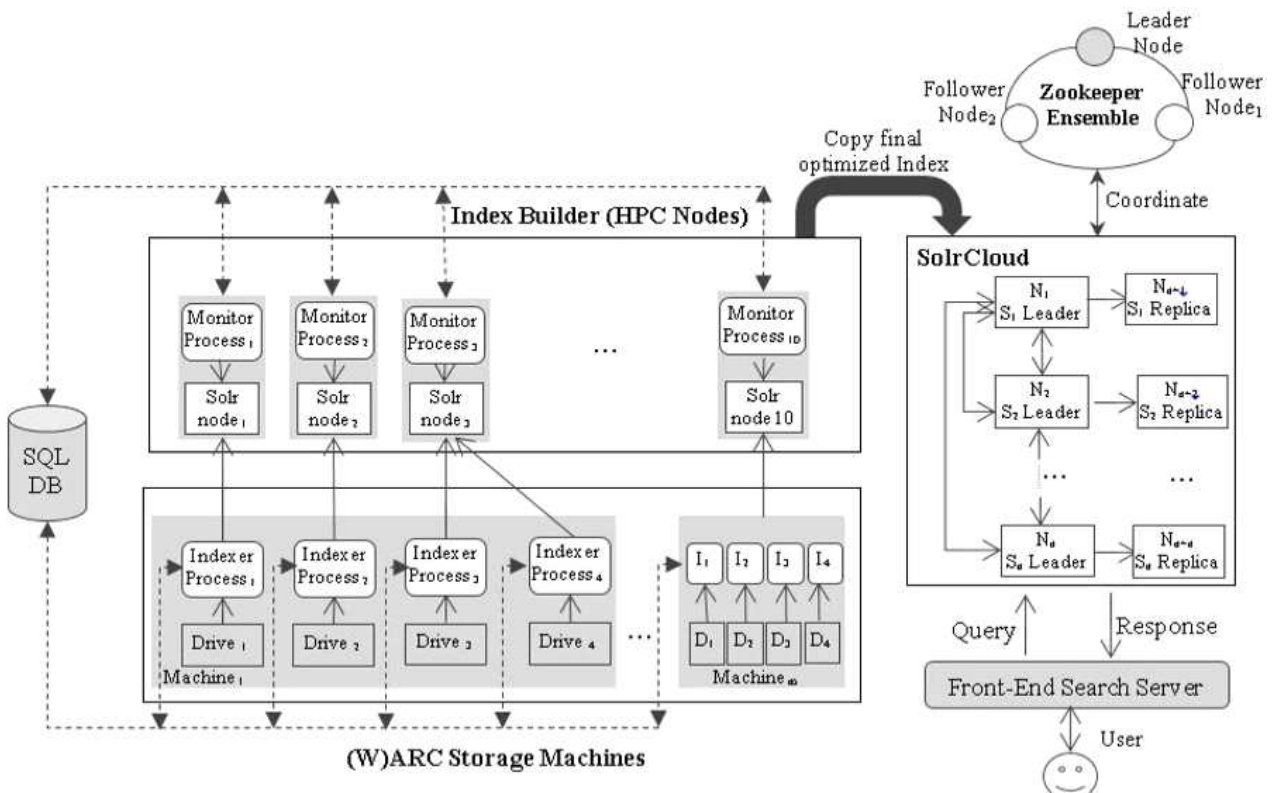
- 발표 : Suzi Szabo(The Dutch National Archives)
- 내용
  - 네덜란드 국립 아카이브는 정부의 웹사이트 보다 잘 아카이빙하기 위해 최근 <웹 아카이브 가이드>를 발행하여 정부기관에 제공하였음.
  - 공공기록법에 따르면 네덜란드 정부기관과 공공 아카이브는 웹사이트를 포함하여 그들이 생산한 기록을 아카이브하는 책임을 공유함.
  - 가이드는 정부기관을 위한 웹사이트 아카이빙의 책임, 프로세스 및 결과에 중점을 둔 일련의 요구사항, 프로세스에서 모든 단계를 기술하는 로드맵 등을 제공하여 웹정보의 지속가능한 접근성 보장에 도움이 됨.
  - 향후 대화형 웹사이트 및 소셜미디어에도 적용할 수 있도록 가이드의 범위를 확장할 예정

### □ 제한된 자원을 사용하여 대규모 웹 아카이브 콘텐츠를 색인하고 검색하는 워크플로우

- 발표 : Sara Elshobaky, Youssef Eldakar(Bibliotheca Alexandrina)
- 내용
  - 알렉산드리아 도서관에 1PB 이상의 웹 콘텐츠를 아카이빙 했으나 full-text 검색 기능이 없으며 오직 URL 검색만 가능
  - 일반적으로 full-text 검색 기능 구현을 위해서는 강력한 시스템이 필요하나 알렉산드리아 도서관(BA)에서는 구형 고성능 컴퓨팅 클러스터를 사용하여 full-text 검색을 지원하는 시스템을 구축한 사례 소개



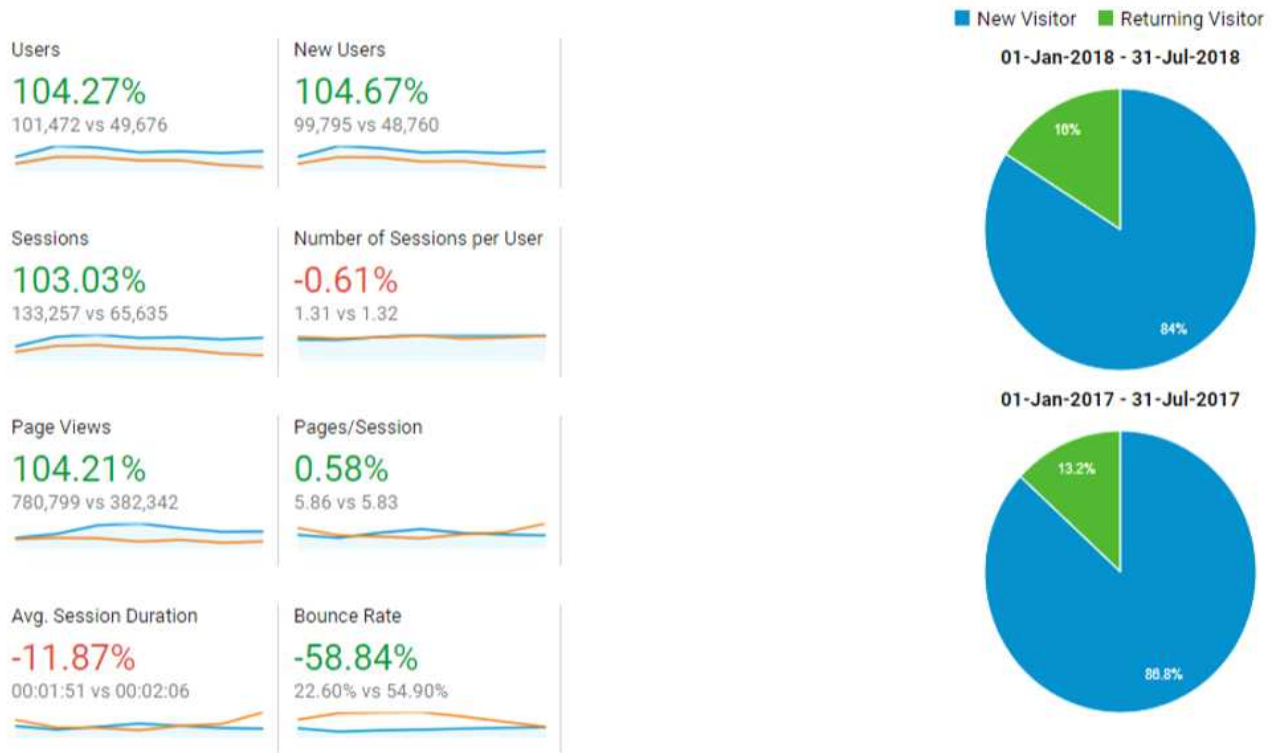
## <시스템 구조>



### □ 포르투갈 웹아카이브(Arquivo.pt) 사이트 소개

- 발표 : Joao Gomes(Arquivo.pt, FCCN-FCT)
- 내용
  - 연구원, 학생, 언론인들을 대상으로 Arquivo.pt에 대한 홍보 및 교육 활동을 실시하여 Arquivo.pt에 대한 이용자 인식 및 사용이 크게 향상된 사례를 소개
  - 주요 홍보 활동 : 언론인 교육, 디지털 마케팅 캠페인, 연구원 보조금 지급, 이벤트, 모범 사례에 대한 비디오 제작, 보도 자료 배포 및 미디어 파트너십 구축

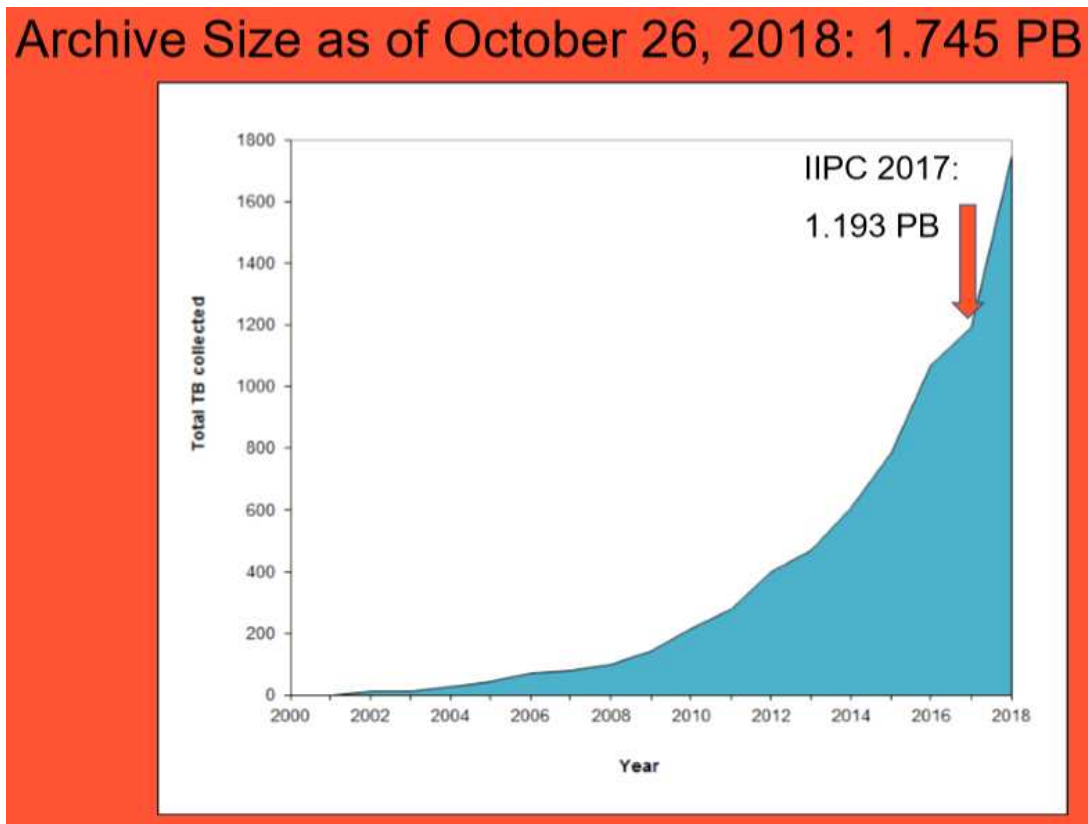
## <2017년과 2018년 사이트 이용현황 비교>



### □ 미국의회도서관의 웹아카이브 진행상황

- 발표 : Abbie Grothe, Grace Thomas(Library of Congress)
- 내용
  - 조직 : Library Services
    - > Office of the Chief Operating Officer for Library Services
    - > Digital Collections Management & Services (DCMS)
    - > Digital Content Management Section (DCM)
    - > Web Archiving Team (팀원: 5명)
  - 컬렉션 : 이전부터 구축해 온 35개 컬렉션을 지속적으로 수집하는 한편, 2018년에 미국 정부 웹사이트, 미국 선거, 젠더 연구, 미국 뮤직 산업 등 다양한 주제에 대한 19개의 새로운 컬렉션 구축 시작
  - LC 내부에서 웹 아카이빙에 대한 관심 증가로 2018년 4차례 회의 개최
  - 신규 발행 컬렉션 : Web Culture, Webcomics, Science Blogs, Foreign Law
  - 웹 아카이브 데이터 세트 다운로드 서비스 실시

- 연도별 수집 현황



## 5. 5일차 주요내용(워크숍)

### □ [워크숍] WARCs로 무엇을 할 수 있을까?(WARCs:아카이브 Unleashed Toolkit)

- 워크숍 코디네이터 : Andrew N. Jackson, Ian Milligan, Olga Holownia
- 내용
  - 원문 색인(full-text indexing) 및 웹 아카이브 자료 분석을 위한 다양한 도구 소개
  - 원문 검색 및 시각화를 위해 웹 아카이브 탐색 색인 시스템, Shine 및 Warclight 사용자 인터페이스 기반으로 진행
  - Archives Unleashed Toolkit, Archives Unleashed Cloud 검토
  - 샘플 데이터를 통해 웹 아카이브 내 텍스트 추출 방법 등 소개

□ **[워크숍] 커뮤니티 협업을 통한 윤리적 소셜 미디어 아카이빙**

- 워크숍 코디네이터 : Jessica Moran, Matariki Williams, Bergis Jules, Edward Summers, Alexandra Dolan-Mescal 및 Francis Kayiwa
- 내용
  - 커뮤니티 활동가, 아키비스트, 사서, 학자, 개발자 및 디자이너 등이 모여 더 풍부하고 비 억압적인 웹 아카이브를 만들 수 있는 방법 논의
  - Documenting the Now 프로젝트 및 다른 소셜미디어 웹 아카이빙 도구 소개

□ **[워크숍] 연구원은 무엇을 원하는가? 검색 및 접근을 위한 Crowdsourcing 요구 사항**

- 워크숍 코디네이터 : Andrea Goethals, Ben O'Brien
- 내용
  - 연구원, 학생, 교사, 일반 대중에게 유용한 웹 아카이브 컬렉션을 만드는 방법, 기술 인프라, 기능, 메타데이터 또는 지원 측면에서 제공해야 하는 것, 다른 이용자 그룹을 위한 요구사항, 웹 아카이브의 창의적인 사용 촉진 방법 등에 대해 논의

# V. 관련 사진



개회(IIPC 운영보고)



컨퍼런스



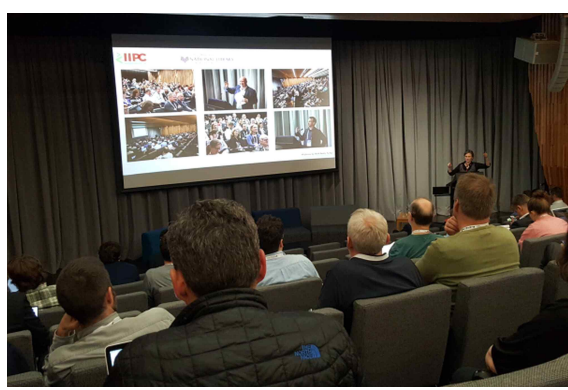
컨퍼런스



컨퍼런스



교류시간



폐회

## VI. 종합의견

---

### □ 웹 아카이빙의 현재

- Webrecorder, Cobweb, Memento, Web Curator Tool 등과 같은 커뮤니티 툴이 지속적으로 업그레이드되고 있고,
- 스트리밍 미디어 수집을 위한 새로운 크롤러 개발, 대량의 웹사이트를 효율적으로 크롤링하는 방법 연구, 소셜미디어 수집을 위한 다양한 시도 등이 진행되고 있으므로 추진경과에 대한 지속적인 관심 필요
- 일반 이용자 및 연구자들의 웹 아카이빙 검색 및 이용을 돕기 위한 웹사이트 인터페이스 연구 및 개발이 활발하게 진행되고 있음.

### □ 국가별 웹 아카이빙 전략

- 납본법과 개인정보보호법의 테두리 안에서 웹사이트를 수집하고 서비스 하는 영국의 웹 아카이브 서비스 전략, 효과적인 웹 아카이빙을 위해 정부기관에 웹사이트에 대한 가이드라인을 제공하는 네덜란드 국립도서관 등 국가 단위의 프로젝트 사례들을 참고할 만함

### □ 기타

- 컨퍼런스에서 엔지니어, 프로그래머 등의 시스템 개발 사례 발표가 많은 부분을 차지하고 있음. 시스템 구축 및 기능 개선 사례에 대한 정확한 이해 및 우수 사례 벤치마킹을 위해 우리 도서관 시스템 담당자의 컨퍼런스 참석이 필요함

※ 선물수령 관련 : 선물수령 여부 :  예.  아니오.