데이터수집 목적 별 크롤링 활성화 제언

Suggestion to utilizing crawling methods by purpose of data collection

김휘강 (고려대학교 정보보호대학원)

차 례

■ keyword : | crawling, intellectual property |

1. 서론

크롤링 (Crawling) 은 원격에 위치한 리소스 (원격 서버와 데이터 통신을 통해 엑세스 가능한 웹페이지, 이미지, 텍스트 등)를 범용적인 TCP/IP 통신, 특히 HTTP(S) 통신을 통해서 가져오는 기법을 의미한다.

크롤링 (Crawling) 은 인터넷 서비스의 핵심 기능 중 하나인 "검색서비스"에서 쓰이는 대표적인 기술이며, 활용방안에 따라 검색엔진 외에도 스캐닝 (scanning), 스크래핑 (scrapping) 기반의 금융서비스, 전자상거래 상에서의 가격비교, 정보보호 서비스 등 다양한 분야에서 활용되고 있다.

기본적으로 크롤링은 인터넷 상에 불특정 다수에게 접근을 허용하는 서비스, 즉 웹서비스와 같이 개방성을 전제로 한 서비스 상에 존재하는 리소스를 통신 프로토콜을 활용하여 수집하는 방식이므로, 기술적으로는 정상적인 리소스에 대한 요청(예: GET method를 통한 Request) 과 특별히 다르지 않다.

이에, 특별히 크롤링을 제한할 필요는 없지만, 서버 및 네트워크 성능이 충분히 확보가 안된 상태인 경우 크롤링을 통해 자원을 공유한 서버에 부하량을 일부 높일 수 있는 문제가 발생한다거나, 서버 관리자의 부주의한 접근제어 설정으로 인해 원래는 공유할 의도가 없었으나 의도하지 않게 외부에서 데이터가 보이게 되는 문제가 발생할 수 있다.

이러한 문제가 발생했던 일부 사례들로 인해 원치 않는 정보유출이 발생할 수 있다는 의견이 등장했으며, 이에 크롤링 자체에 대해 제약을 강하게 두자는 의견이 있어왔다.

이에, 본고에서는 크롤링의 다양한 사례에 대해 살펴보고, 크롤링의 긍정적인 활용 사례를 제시하여, 크롤링에 대해 국내에 다소 과도하게 자리잡은 부정적인 견해를 불식하고, 크롤링 기술의 활성화 방안을 제시하고자 한다.

2. 법적 기술적 고찰

2.1 법적 고찰

2.1.1 정보통신망법상 판례

크롤링과 관련되어 주목해 볼만한 판례들은 다음과 같다.

(1) O2O 비즈니스 제공 기업 중 하나인 '여OOO'에서 '야OO' 사의 웹페이지를 지속적으로 크롤링 한 뒤 데이터를 축적한 것이 일종의 '무단복제' 로 볼 수 있다고 한 판례가 있다. (서울중앙지방법원 2020.2.11. 선고 2019고단1777 사건)

이에 대한 항소 위에서 예를 들은 '여OOO' 와 '야OO' 의 경우에는 항소심에서 다른 판례가 나온 점이 많은 시사점을 준다고 할 수 있다. 1심에서는 크롤링이 검색 엔진 등에 널리 사용되고 있지만, 일종의 타인의 정보통신망을 무단으로 침입한 것으로 보고 있는 견해가 주 판단근거였던 반면, 항소심에서는 '여OO O'에서 가져간 정보가 번거롭긴 하지만 수동으로 크롤링을 한 것과 같은 종류와 양의 정보를 가져갈 수 있다는 점을 주 근거로 삼아 다르게 판단하였다. '야OO' 측에서 적극적인 API 서버 접속 접근제어를 하지 않은 점 역시 1심의 판단을 바꾸는데 영향을 주었다고 볼 수 있다.

(2) Wikipedia 와 같이 유저들의 참여와 집단지성으로 컨텐츠가 생성되는 다양한 Wiki 기반 웹사이트들이 있다. 국내의 Wiki 유형의 서비스 중 하나인 '리그OO' 위키와 '엔OOO 미러' 간

의 발생했던 법적 분쟁 사례 역시 크롤링과 관련된 대표적인 분쟁이라 할 수 있다.

'엔OOO 미러' 사이트에서 '리그OO' 위키의 자료를 그대로 복제해 광고 이익을 얻은 것이 원글 자체는 UCC (User Created Contents) 특성상 '리그OO'위키 사이트가 컨텐츠 자체에 대한 저자는 아니지만, 이 컨텐츠들을 이용하기 용이하도록 데이터베이스를 구축하고 페이지 구성 설계를 한 것 및 운영 유지관리를 하고 있는 점을 들어 '엔OOO 미러' 사이트가 복제권 및 전송권 침해를 하였다는 판례가 있다. (서울고등법원 2016.12.15. 선고 2015나2074198 사건)

위에서 언급한 두 판례들의 주요한 쟁점은 크롤링 그 자체가 통신망을 침해한 행위로 볼 수 있느냐는 기술적인 관점과, 크롤링을 한 행위의 목적이 부당이익을 얻기 위해 고의성을 가지고 한 것인가로 볼 수 있다.

우선, 두 사례 모두 크롤링이라는 행위 자체가 침투적인 '해킹'에 가까운 행위나 정상적인 인터넷 서비스를 '방해'하는 관점이라기 보다는, 데이터베이스의 상당한 부분을 복제하여 부당한 이익을 얻었는가가 판단에 큰 영향을 끼친 것으로 볼 수 있다.

저작권법상에서 언급된 데이터베이스의 상당한 부분에 대한 복제 및 데이터베이스 저작자의 이익을 부당하게 침해하는 범위가 아니라면, 크롤링을 통한 정보 수집 자체에 대해 위법하다고 단정짓기 어려운 면이 있다.

하지만, 크롤링을 통한 정보 가공을 통해 인터넷 이용자들에게 보다 높은 가치의 정보를 재생산할 수 있음에도 불구하고 위의 대표적인 사례들이 지적재산권, 저작권 침해요소와 관련된 사례이다 보니, 크롤링의 목적 및 활용방안과 무관하게 크롤링 기술 자체에 대해 부정적인 인식을 과도하게 갖게 된 면이 있다.

위에 언급된 대표적인 두 판례들과 더불어, 스크래핑 기술에 대해서도 부정적인 인식이 자리잡고 있다. FinTech 서비스 활성화 초창기에는 금융 정보를 활용할 수 있는 API 가 없었으므로, 크롤링의 한 유형인 스크래핑 방식을 활용하여 산재된 금융사이트 이용 정보를 수집해야 했었다.

국내의 경우 초기 FinTech 서비스들이 등장할 즈음에, 금융기관에서 보안상 및 제도적인 준비가 마련되기 전이었기 때문에 API를 통해 금융계좌정보, 잔고 등의 정보를 외부에 노출하는 것이 상당한 제약이 있었다.

그래서, 서비스에 가입한 이용자에 의해 인증에 필요한 정보를 받은 뒤, 인증된 웹접속 세션을 통해서 (즉, 로그인에 필요한 크레덴셜 정보를 활용하여 다양한 금융사이트들의 잔고 및 카드사들의 포인트 정보를 취합하는 방식) 은행, 증권, 카드사들의 웹사이트 상의 정보를 가져오는 스크래핑 방식을 택했었다.

다만, 이 시점에서는 금융권 역시 새로운 방식의 접근 패턴에 대해 아직 준비가 되어 있지 않은 상태였기 때문에 CPU, 네트워크 트래픽 용량 면에서 가용성 확보가 안 된 서버들에 부하를 유발하는 문제가 있었다.

위와 같은 스크래핑 패턴은 논리적으로는 개별 이용자가 각자 여러 사이트를 방문하여 정보를 취합하는 것을 대행해 주는 것과 다를 바 없으므로 침해나 서비스방해와는 거리가 있으나, 결과적으로는 스크래핑 방식의 FinTech 서비스가 인기를 얻으면서 접속이 급증함에 따라 의도치 않은 서비스 부하를 유발한 것으로 볼 수 있다.

이로 인해 크롤링 또는 크롤링 기반 유사한 데이터 수집 방식에 대해 리소스 제공자 입장에서는 매우 부정적인 인식을 갖게 되는 계기가 된 것 역시 사실이다.

2.2 기술적 고찰

2.2.1 네트워크 접속 관점에서의 고찰

우선, 웹사이트를 크롤링하는 경우를 먼저 살펴보면, 크롤링 자체는 사람에 의한 접속에 비해 웹페이지들에 대한 네비게이션 (사이트 내의 링크들을 클릭하여 이동하는 경로) 탐색 속도가 빠르다는 점 외에는 통신에 사용되는 프로토콜 (예: HTTP, HTTPS) 및 METHOD (예: GET 방식) 은 동일하다.

크롤링이라는 용어가 등장하기 전에도 webzip 이나 wget 과 같은 유틸리티를 이용하여 사이트 컨텐츠를 편리하게 내려받는 것 역시 표준 웹 프로토콜을 이용하였던 예라 할 수 있다.

또한, 보편적으로 많이 사용되는 검색엔진 서비스들이 robot을 이용하여 웹사이트를 수집해 가는 것 역시 모두 동일한 방식이다.

웹서비스 제공자 역시 크롤링이 기술적으로는 일반 이용자들의 접근과 다르지 않기 때문에 다음과 같은 관례를 통해 크롤러를 식별하고 제어를 할 수 있었다.

첫번째로 웹서비스 제공자가 방문자 분석을 하거나, 접근을 제한하기를 희망할 때 판단 근거로 쓸 수 있도록, 웹서비스 제공자에게 지금의 접속이 사람에 의한 것이 아닌, 크롤링 프로그램에 의한 접속임을 알려주기 위해, 접속을 하는 크롤러의 user-agent 정보를 범용적인 웹브라우저 (예: Google Chrome, Microsoft IE) 와 다른 정보를 설정하여 제공하였다.

두 번째로, 웹서버 상에 robots.txt에서 정의된 설정을 보고 특정 디렉토리는 접근할 수 있다 할지라도 크롤링을 안하는 것이 관례로 되어 있었다.

이 정보를 토대로 웹서버에서는 동일 IP address 와 동일 use

r-agent 에서의 초당 접속 수를 제한하거나 CAPTCHA 인증을 이용하여 사람과 robot 의 접속을 구분하는 것이 가능했다.

즉, 웹서비스가 최초 인터넷에 등장할 때의 개념 자체가 불특정 다수의 모든 이용자가 어디서나 자유로이 리소스를 액세스하도록 설계되어 있기 때문에, 명시적으로 웹서비스 제공자 측에서 특별한 접근제어 및 보호조치를 하지 않았다면, 웹 서비스 상에 업로드된 리소스들은 기본적으로는 어떠한 방식으로든 제약없이 엑세스 가능한 것으로 보아야 한다.

2.2.2 서버 관점에서의 고찰

통신포트 80, 또는 443을 이용하여 서비스 되는 웹서비스 외에도 다양한 인터넷상의 서비스가 1~1024 번 통신포트 상에서 통칭 well-known service 라는 이름으로 서비스되고 있다.
well-known service 의 예로는, 25 번 통신포트에서 서비스되는 SMTP 프로토콜 기반의 이메일 서비스, 23 번 통신포트에서 서비스 되는 telnet 로그인 서비스, 1433번/1434 통신포트에서 서비스 되는 MS-SQL 서비스 등을 들 수 있다.

인터넷에서 통신을 하기 위해서는 우선 해당 통신포트가 열려있는지 접속을 맺고, 해당 통신 프로토콜에 정의된 대로, 초기 핸드쉐이킹 메시지를 보내게 된다. 이에 대한 응답으로 서비스 버전 및 통신프로토콜 명령어가 올바르게 동작하는지를 판단하여, 통신을 지속할 것인지 종료할 것인지를 정하게 된다.

즉, 통신포트가 열려 있다라는 것은 기본적으로 들어오는 접속을 허용한다는 것을 의미하며, 이 과정에서 서버는 접속자들에게 자연스럽게 서버의 프로덕트 및 버전 정보 등 부가 정보를 공개적으로 제공하는 것으로 간주할 수 있다.

이와 같이, 인터넷 상에 서비스 되는 모든 통신들은 개방성을 기반으로 구현이 된 것이므로, 명시적으로 접근제어를 설정하고 있지 않다면 인터넷 상의 불특정 다수가 해당 통신포트에 접속하여 서비스 정보를 알도록 한 것으로 볼 수 있다.

마치 실 세계에서 상점이 간판을 외부에 내걸고, 문을 열어두어 영업 중 임을 알리는 것과 유사한데, 행인이 영업 중 인지를 확인하기 위해 간판을 상세히 읽고, 문이 열려 있는지 확인하기 위해 상점에 접근한 것 자체가 허용되는 것과 유사하다고 보아야 한다.

다만, 상가 내 상점들을 빠르게 달리면서 훑어보거나 문 손잡이를 잡았다가 상점 내로 들어오지 않고 다른 상점으로 이동하는 행위를 방어적으로 해석하면 공격을 하기 위한 전단계인 스캐닝 (scanning) 으로 간주할 수도 있다.

하지만, 현실적으로 스캐닝을 했다는 행위만으로 불법으로 처벌을 하고 있지는 않은 현실을 감안해 볼 때, 공개된 통신포트를 통해 노출된 정보에 대해 목적성과 관계없이 스캐닝, 스크래핑, 또는 크롤링을 했는지를 제약한다면 메타검색에 의한 가격비교, FinTech, 보안 관련 신규서비스 기술발전을 저해하는 요인이 될 수 있다.

예를 들어 10년전이라면, 단일 IP address에서 분당 10 건의 접속만으로도 서버에 부하를 줄 수 있으므로 제약을 해야 한다는 의견이 있었다면, 지금은 클라우드 기술 및 웹서버 성능의 향상으로 인해 초당 100건의 접속 자체도 제약을 하기에도 애매한 숫자가 되었으며, 기업 등 NAT 환경에서 대표 IP address 로 접속하는 이용자들의 동시접속을 제약하는 경우가 발생할 수 있으므로, 서비스 제공자 관점에서 크롤링 접속을 제한하는 것은 기준이 기술적으로 매우 모호함을 알 수 있다.

2.2.3 컨텐츠 보안 관점에서의 고찰

2.2.2 절에서 살펴본 것처럼 인터넷의 서비스 자체가 태생적으로 개방성을 전제로 하고, 기 오픈되어 있는 서비스는 공개된 컨텐츠임을 고려해 볼 때, 접근제어의 책임이 정보를 보호하려는 주체에 달려 있는 것이 자명하다고 할 수 있다.
공개된 자료이기는 하지만, 특정한 대역의 IP address 상에서만 접근을 하기를 원하거나, 회원 중에서도 결제를 완료하였거나 로그인이 완료된 이용자에게만 액세스를 하도록 제약하고 싶은 컨텐츠가 있다면 서비스 제공자가 이에 대한 기술적 보호조치를 취해야 한다.

네트워크 상에서 명시적으로 접근제어를 하는 방법으로는 Firewall을 통한 IP address 기반의 접근제어 역시 가능하고, 데이터베이스 상에 적재된 컨텐츠의 경우에는 DBMS 의 접근제어 기능을 이용하여, 접속 계정의 종류, 등급에 따라 제어를 하는 것이 가능하다.

이 외에도 웹상으로 공개된 텍스트 및 이미지 기반 컨텐츠 자체에 대해서도 기본적인 관리적 보호조치를 해두는 것이 지적재산권을 보호하는데 필요하다고 할 수 있다.
예를 들어, 웹페이지 상에 공개된 저작물들에 대해, CC (Creative Commons) [1] 표기를 통해 저작권 레벨을 알리고, 이미지의 불법적인 재활용을 예방하기 위해 워터마크 삽입을 하여 텍스트 및 이미지 기반의 컨텐츠에 대해 지적재산을 표기하는 방법이 있다.

이 외에도 PDF 와 같은 특정 포맷의 문서형 자료인 경우 문서 내의 보호기능을 활용하는 방식 등 웹상에서 제공할 컨텐츠에 대한 보호조치를 적용하는 방식 역시 가능하다 [2].

요컨대, 크롤링으로 인한 잠재적인 피해에 대비하기 위하여 “some allow, all deny by default” 방식을 취하기 보다는, 웹을

통해 서비스 되는 리소스를 보호하기 위해서는 보호해야 하는 자산 (asset) 인지를 서비스 제공자 측에서 식별을 하여 보호조치를 하고, 기본적으로는 개방성을 유지하는 "some deny, all allow by default" 방식이 웹서비스의 취지에 부합한다고 볼 수 있다.

3. 크롤링 기술의 긍정적인 활용 사례

3.1 검색엔진 서비스

크롤링은 검색엔진에서 전통적으로 활용되어 왔다. 검색엔진은 사실상 웹으로 대표되는 인터넷 서비스의 활성화를 가져온 대표적인 기술로서, 사용자가 엑세스하고자 하는 컨텐츠, 서비스가 어느 URL에서 제공되고 있는지를 연결하는 Human-Machine Interface 역할을 담당하고 있다고 보아도 과언이 아니다.

단순한 컨텐츠 리소스의 location (URL) 정보만을 제공해주는 것 외에도 사전에 크롤링을 통해 수집한 정보를 토대로 쇼핑몰 상품, 여행 상품 등 온라인상에서 구매 가능한 모든 종류의 서비스 및 상품에 대한 가격비교와 같은 이용자에게 보다 높은 가치를 제공해주는 서비스를 창출하는 것이 가능하다.

3.2 보안 서비스

3.2.1 온라인게임 지적재산권 침해 사이트 탐지

온라인게임에서는 상용 게임서버를 불법적으로 해킹을 통해 서버 프로그램을 탈취하거나 역공학을 통해 서버를 개작하여 만든 뒤, 불법적으로 서비스하는 "사설서버" 들이 존재한다.

서은비 등의 연구 [3]와 유창석, 김휘강 의 연구 [4], 김기범의 연구 [5]에 의하면 불법사설서버로 인한 연간 총 피해규모는 2017년을 기준으로 2조 4,385억으로 예상되는 등 , 사설서버는 온라인게임 산업의 발전을 저해하는 주요 원인이라 할 수 있다.

사설서버는 명백한 지적재산 침해행위로서 저작권법 상에서 제 101조의4 (프로그램코드역분석)에서 "호환 목적 외의 다른 목적을 위하여 이용하거나 제3자에게 제공하는 경우" 및 게임산업진흥에 관한 법률 상에서 제 32조의 9 (불법게임물 등의 유통금지 등)에서 "게임물 관련사업자가 제공 또는 승인하지 아니한 게임물을 제작, 배급, 제공 또는 알선하는 행위" 에 해당된다고 볼 수 있다.

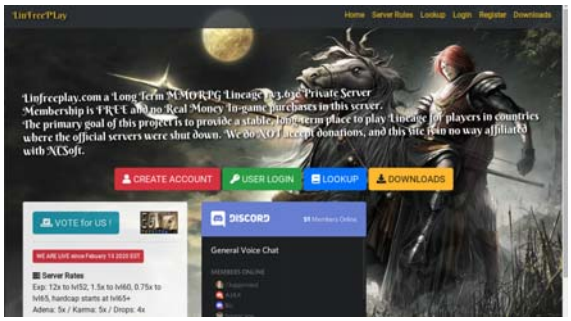
[그림1] 은 크롤링 기술을 활용하여 전 세계에 존재하는 사설서버를 탐지하는 AI Spera 社의 pitection 서비스의 예이다.

이 서비스에서는 크롤링 기술을 통해 공식 게임사이트가 아닌 사이트들에서 동일한 이름의 게임명이 검출되는지를 확인한 뒤, 머신러닝을 응용한 유사도 검색을 통해 해당 사이트에 개작된 게임 및 유사 이미지가 업로드 되어 있는지, 자연어처리를 활용하여 텍스트 분석을 하여 사설게임서비스가 제공되고 있는지를 판단하게 된다.

과거에는 각 게임사의 사설서버 담당자들이 수작업 검색을 통해 대응하거나, 이용자들의 자발적인 신고에 의존하였기 때문에 검출량이나 범위 면에서 한계가 있어왔다.

위 서비스는 이 과정을 자동화 하여 사설서버를 손쉽게 식별하게 된 예로서, 크롤링 기술을 통해 편익을 제공한 긍정적인 사례라 할 수 있다.

▶▶ 그림 1. AI Spera사의 Pitection 서비스에서 탐지된 사설서버



▶▶ 그림 2. 탐지된 사설서버를 스크린캡쳐를 통해 증거보존한 예

3.2.2 사이버위협정보 (Cyber Threat Intelligence) 서비스

인터넷의 주요 서버들에 대해 자주 해킹시도를 하는 IP address 들을 알아낼 수 있다거나, 취약한 버전의 소프트웨어를 운영하고 있어 해킹당하기 쉬운 잠재적인 victim 서버들에 대한 정보, 그리고 피싱 사이트 등 악성 도메인 정보를 사전에 파악할 수 있다면 미리 해당 서버들의 IP address 및 도메인으로부터의 접속을 차단할 수 있을 것이다.

이러한 정보를 사이버위협정보 (Cyber Threat Intelligence; CTI) 라고 부르며, 사이버위협정보를 활용하여 유입되는 공격을 예방차원에서 사전에 차단할 수 있다.

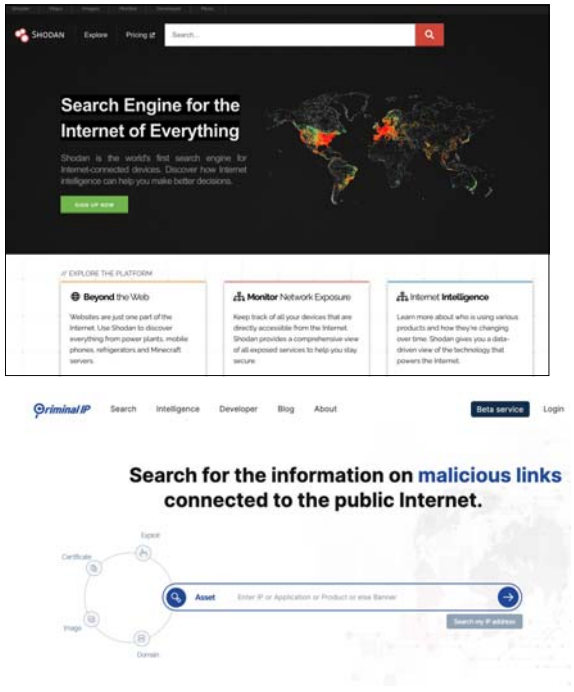
크롤링은 이러한 사이버위협정보를 획득하는 데에 핵심적인

기술이다. 사이버위협정보 제공사들은 크롤링을 통해서 전세계 인터넷 상의 IP address, 도메인, 취약점 정보, 각 IP address 및 도메인에서 서비스 중인 정보들에 대해 수집하고 이를 데이터베이스에 저장한다. 그 뒤, 크롤러에 의해 수집된 데이터를 분석하여 악성코드나 악성 URL 이 삽입되어 있는 웹페이지 및 IP address, 도메인 정보를 식별한 뒤, 추출된 정보를 통신사/ISP 및 금융권 등에 제공하여 잠재적으로 발생할 수 있는 피해를 조기에 방지할 수 있도록 한다.

최초에 shodan.io 와 같은 사이버위협정보 서비스가 출현하던 시점만 하더라도, 정보수집을 하는 과정에서 활용되는 크롤링이 통신망 침해의 일환으로 보는 것이 맞지 않느냐는 견해도 지속적으로 있어왔다.

하지만, 크롤링의 목적이 보안서비스로서 인터넷 이용자들과 많은 기업들을 보호하는데 도움이 된다는 긍정적인 면이 크기 때문에, 현재는 크롤링과 관련된 부정적인 이미지나 논란은 거의 사라진 상태이다.

오히려 시장에서 사이버위협정보 서비스의 중요성을 인식하게 되어, shodan, security scorecard, domiantools, AI Spera 등 많은 국내외 유수의 전문 기업들이 크롤링 기반으로 획득한 정보를 토대로 사이버위협정보를 서비스하고 있다.



▶▶ 그림 3. 위협정보서비스 사이트 예시 (위: shodan, 아래: Criminal IP)

4. 데이터수집 목적 별 크롤링 활성화 제언

4.1 서비스 목적 별 크롤링 활성화 방안

앞서 설명한 것처럼, 네트워크, 서버 상에서는 지금 유입된 접속이 선의의 목적의 사이트 엑세스인지 여부를 판단하기에는 기술적인 모호성이 존재한다.

이는 포트스캐닝이 보안상 침해가 아니냐는 관련 논의와도 공통점이 많다고 할 수 있다.

즉, 해커일 수도, 아닐 수도 있는 신원 불명의 인터넷의 이용자가 단순히 포트스캐닝을 했다는 이유만으로, 이러한 포트스캐닝 행위는 해킹의 전단계로서 미수에 그친 범죄 또는 계획범죄의 일환이었다고 단정할 수 없는 것처럼, 크롤링으로 판단되는 접속이 유입되었다고 해서 이 행위가 지적재산권 유출 행위의 전단계라고 판단할 수 없다고 할 수 있다.

더불어 과거 판례들에서 크롤링을 통한 침해 관련 판단시 자주 언급되는 개념 중에서, 웹사이트에 “과도한 접근”, “데이터베이스에 상당한 부분에 대한 복제” 라는 것이 있다.

하지만, “과도한”, “상당한” 이라는 것은 모호한 면이 높다. 과거에는 과도한 접속으로 여겨졌던 것이 시스템 및 네트워크 기술이 발전함에 따라 손쉽게 처리 가능한 동시접속이 되어 버린 것처럼 기준값으로 삼을 수 있는 임계치를 정량적으로 산정하는 것 역시 무리가 있다고 할 수 있다.

요컨대, 크롤링의 최종적인 의도가 어떤 것이었는지 파악하기 어려우며, 크롤링을 기반으로 한 유용하고 편익이 높은 서비스 역시 혼재 되어 있으므로 단순히 유입되는 정보만을 가지고 의도성을 판단하는 것은 기술적으로도 법적으로도 한계가 있다고 볼 수 있다.

본고에서는, 크롤링 관련 분쟁이 발생 시 (예: 크롤링이 에 의해 서버의 가용성 저하 등), 일괄적으로 침해적인 행위로 보아 제약을 하기 보다는, 크롤링은 유익한 서비스에도 보편적으로 활용되는 기술이므로 서비스에 문제를 일으키지 않으면서 보다 활성화 할 수 있는 기술적인 방안을 강구해 나가는 것이 필요하다고 판단된다.

5. 결론

현재 크롤링이 저작권 침해관련하여 사용되었던 전례를 들어, 크롤링에 대해 제약사항이 과도하게 적용하고 부정적인 면이 강조되어 있는 상황이라 생각된다.

본고에서는 크롤링과 관련된 과거의 법적인 분쟁 사례 및 유익한 용도로 활용되는 사례를 제시하여 크롤링에 대해 보다 기술중립적인 시각으로 바라볼 것을 제안하고자 한다.

서비스 제공자는 보호가 필요한 리소스에 대해서는 명확한 기

술적 보호조치를 하고, 기본적으로는 인터넷 서비스 상에 운영 중인 모든 서비스는 개방성을 기본으로 한다면, 향후 크롤링 기반의 다양한 유익한 서비스들의 출현을 촉진 시킬 수 있을 것으로 판단된다.

또한 웹서버에 직접적인 크롤링 방식이 서버 및 네트워크 상에 부하를 일으킬 수 있는 단점이 있으므로, 웹 서버 외에도 별도의 API 서버를 통한 데이터 접근 방안 역시 제공을 하는 것이 접근성을 다양화 한다는 점에서 권장된다고 할 수 있다. 즉, 기본적인 크롤링은 모두 허용하되 대량의 크롤링의 경우에는 유료 API를 제공하거나, 사이트 내의 컨텐츠에 대해 데이터베이스를 유료로 판매하여 부작용 역시 최소화 하면서 수익성도 높이는 방식들이 자리 잡기를 기대해 본다.

참고문헌

[1] Creative Commons, https://creativecommons.org
[2] Security Cofiguration Guide for Acrobat, https://www.adobe.com/devnet-docs/acrobatetk/tools/AppSec/index.html
[3] 서은비, 김지홍, 강태운, 유창석, 김휘강.(2017).불법사설서버 현황 및 대응방안.정보보호학회지,27(4),27-35.
[4] 유창석, 김휘강.(2017).네트워크 기반 정보재의 불법 사설서버로 인한 피해규모 추정에 대한 연구.한국혁신학회지,12(3),67-82.
[5] 김기범.(2019).온라인게임 사설서버의 범죄실태와 형사정책 개선방안.법학연구,19(2),29-52.

저자소개

● 김 휘 강(Huy Kang Kim)

■ 1998년 2월 : KAIST 산업경영학과 학사
■ 2000년 2월 : KAIST 산업공학과 석사
■ 2009년 2월 : KAIST 산업및시스템공학과 박사
■ 2010년 3월 ~ 현재 : 고려대학교 정보보호대학원 교수

〈관심분야〉 data-driven security, 사이버위협정보, 자동차 보안