

방송 미디어를 위한 AI 기술 - 2

- 다양한 방송 미디어 응용과 AI 기술 적용 방식

글. 오주현 KBS 미디어기술연구소

연재 목차

1. AI 기술의 개요와 방송 미디어 응용
2. 다양한 방송 미디어 응용과 AI 기술 적용 방식

지난 1부에서는 방송 미디어를 위한 AI 기술 개요에 대해 알아보고, AI의 방송 미디어 활용 분야로 뉴스 프레젠테이션, 품질관리, 메타데이터와 검색에 대해 살펴보았다. 이번 2부에서는 다른 몇몇 추가 활용 분야와 AI 도입 방식에 대해 논의하고자 한다.

자동 제작

AI가 사람들의 관심을 끌면서 AI가 영화의 예고편을 제작했다는 등의 뉴스를 아마 접한 적이 있을 것이다. 예를 들면 IBM의 왓슨(Watson)은 100여 편의 공포영화 트레일러를 학습하여 공포영화(트레일러)의 문법을 익힌 후, 전체 내용이 주어진 공포영화 'Morgan'에서 가장 인상적인 장면 10여 개를 6분 만에 추출한 뒤 편집했다. 음악 트랙 등 필요한 부분은 사람의 손을 거쳤다. 마찬가지로 렉서스도 AI가 작성했다는 광고 대본으로 자동차 광고를 만들어 홍보하였다. 온라인에서는 한 소프트웨어 키보드 업체가 해리포터를 학습시킨 후 단어 예측을 통해 해리포터 팬픽션(fan fiction)을 만들어내어 그 엉뚱한 내용이 레딧(reddit) 등에서 화제가 되기도 하였다.

사실 이러한 기술은 AI 관련 뉴스로는 굉장히 각광받을 수 있는 내용이지만, 실제 내용을 검토해보면 AI가 실제 영상을 만들어냈다고보다는 대본을 작성하였을 뿐이거나 편집을 위한 장면을 골라냈을 뿐인 경우가 많다. 앞으로 빠른 제작시간을 필요로 하는 일부 응용 위주로 검토할 수 있을 것이지만, 아직 제작자의 의도나 아이디어를 반영하기 어렵고 창작의 영역이라는 점을 감안하면 실제 도입에는 시간이 걸릴 것으로 생각된다.



그림 1. 각각 (좌)왓슨이 제작한 영화 트레일러, (우)AI가 대본을 쓴 광고의 한 장면



이와 같이 콘텐츠를 직접 제작하는 것과는 조금 다른 응용으로서, BBC는 IBC 2018을 통해 최소한의 제작 인원으로 준 실시간(near-live) 이벤트 중계를 할 수 있는 ‘Ed’ 시스템을 발표하였다. 이 시스템은 방송 스튜디오의 일반 프로그램 제작 인원을 AI로 대체하기보다는, 다수 동시 현장 중계를 하기 위한 목적이 크다. 예를 들면 수백 개의 공연장에서 공연이 이루어지는 글래스톤베리나 에든버러 프린지 등의 음악 페스티벌에서는 기존 중계차 등을 이용해서는 극히 일부 공연만이 중계될 수 있는데, 이 같은 자동화 기술을 이용하면 더 많은 이벤트를 동시 제작 송출할 수 있다. Ed는 3대 정도의 4K 카메라로부터 입력된 4K 영상을 미리 정해진 가이드라인을 따라 크롭(crop)하여 적절한 크기의 HD 프레임을 제작한다. [그림 2]와 같은 장면에서 Ed는 가끔 등장인물이 일부만 보이게 크롭하는 등 제작 가이드라인을 지키지 못했고, 거의 항상 자연스러운 샷을 잡는 전문 카메라맨에는 미치지 못했지만 많은 경우 ‘괜찮은’ 프레임을 생성하여 사용 가능성을 입증하였다.



그림 2. Ed 시스템. (좌)크롭의 예, (우)실제 Ed와 전문가가 각각 생성 및 촬영한 영상

지금 KBS에서는 일부 ‘보이는 라디오’ 서비스를 위해 360도 VR 서비스를 적용하기 시작했는데, Ed와 같은 시스템은 이러한 응용에도 대체 또는 보완적으로 사용할 수 있을 것으로 기대된다.

미디어 포렌식



그림 3. Deepfake를 이용해 영화 ‘터미네이터 2’에서 아널드 슈워제네거를 실버스터 스탈론으로 교체한 영상

흔히 ‘Seeing is believing.’이라고 한다. 우리말로는 백문이 불여일견이다. 그러나 최근의 ‘Deepfake’ 비디오를 보면 더 이상 보인다고 해서 믿을 수 없는 세상이 되었다. 오바마와 트럼프 등 유명인 합성으로 알려진 Deepfake는 최근 [그림 3]에서 보듯이 할리우드 영화의 배우를 꽤 그럴듯하게 바꿔 치는 데 사용되고 있고, 이런 영상들은 이제 유튜브에서 흔히 볼 수 있게 되었다.

만약 Deepfake 때문에 우리가 뉴스 영상을 보고도 더 이상 믿지 못한다면 어떻게 될까? 또는 실제로 발생한 사건을 가짜뉴스로 치부하고 넘어가게 된다면? 지금도 홀로코스트와 달 착륙 등을 부정하고 음모론을 제기하는 사람은 무시 못 할 정도로 많다. 여기에 Deepfake를 이용한 가짜뉴스가 횡행하고 일일이 팩트 체크에 매달려야 한다면 그에 따르는 사회적 비용은 상상을 초월할 것이다. Deepfake는 뉴스뿐 아니라 연예인과 일반인 등을 포르노물에 합성하는 데 사용되어 이미 큰 사회적 물의를 빚고 있다. 심지어 최근에는 여성이 등장하는 이미지에서 옷을 제거해주는 ‘Deep nude’가 출시되었다가 많은 비난을 받고 앱 스토어에서 철수한 바 있다.

‘미디어 포렌식’은 이러한 미디어 조작을 밝혀내는 일이라 할 수 있다. Adobe는 UC 버클리 대학과 함께 포토샵의 Face-Aware Liquify 도구를 이용해서 조작된 사진들을 분석하고 조작 여부를 검출하는 연구를 수행하였다[그림 4]. 이 연구에서 사람은 조작된 사진의 53%만을 찾아냈지만 AI는 99%를 검출했다고 한다. 아울러 AI는 조작된 사진의 원본을 추정하기도 했다.

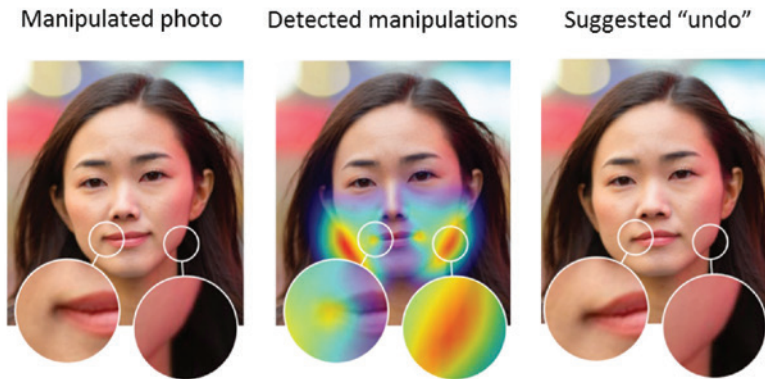


그림 4. Adobe의 조작 사진 검출

Deepfake는 저널리즘을 넘어서 우리 사회 전반에 큰 위협이 되고 있다. Deepfake 기술 발전에 따라 이를 검출하기 위한 AI도 함께 발전하고 있어, 당분간 일종의 군비 경쟁이 이루어질 것으로 보인다. 관련 내용이 CNN의 “When seeing is no longer believing: Inside the Pentagon’s race against deepfake videos” 기사¹⁾에 잘 정리되어 있으니 참고하기 바란다.

광고

최근 디지털 광고에 비해 급격히 비중이 축소되고 있는 TV 광고를 살리기 위한 일환으로, 영국의 Channel 4는 ‘Contextual Moments’라는 새로운 광고판매 확대 기법을 제안하였다. ‘문맥적 순간’으로 번역할 수 있을 Contextual Moments가 제안하는 것은 간단히 예를 들어 설명하자면 이런 것이다. 요리나 식재료 등에 관한 광고 캠페인이 있다고 가정하자(온라인 광고가 아닌 리니어 TV 광고에 관한 이야기다). 광고 효과를 극대화하기 위해 이러한 광고는 예를 들면 ‘The Great British Bake Off’와 같은 요리 전문 프로그램의 전후에 배치하는 것이 선호된다. 만약 이러한 광고를 요리 전문 프로그램뿐만 아니라 요리 장면이 등장하는 일반 드라마와 예능프로그램 등 전후에도 배치할 수 있게 된다면 방송사로서는 광고 스폿의 증대와 그에 따른 수익을 얻을 수 있을 것이다. 일반 프로그램의 영상/음성 분석을 통해 광고 관련 내용(여기에서는 요리)이 포함되었는지 파악하기 위해 AI가 사용된다. AI를 훈련시키기 위해 사람이 ‘여행’, ‘음식’, ‘로맨스’ 등 문맥 광고와 연관될 수 있는 키워드를 태깅한다. 영상으로부터 브랜드와 관련되는 광범위한 객체 및 그 부각 정도를 분석하고, 음성 및 스크립트로부터 주제 및 대화의 의미를 분석하여 각 순간의 긍정성(positivity)을 도출한다. 관련성(문맥)만이 아니라 긍정성을 함께 도출하는 이유는, 음식이 등장하지만 맛이 없거나 상해서 버리는 장면이라든지 자동차가 등장하지만 사고로 부서지거나 불타거나 하는 장면을 시청한 후에 음식/자동차 관련 광고를 시청하게 되면 오히려 부정적인 광고 효과를 가져올 것이기 때문이다. Channel 4가 광고 연구 기관과 함께 Contextual Moments 광고 효과를 분석한 결과, 시청자가 광고를 기억하고 있는지를 나타내는 광고 리콜(ad recall)이 일반 광고에 비해 2배나 상승한 것으로 나타났다. 2018년 여름 발표한 이후 1년이 되어 가는 Contextual Moments의 그 이후 상용화나 실제 적용 사례 등은 아직 출판되지 않았지만, 앞으로 리니어 TV 광고의 생존을 위해 시사하는 바가 크다.

1) <https://edition.cnn.com/interactive/2019/01/business/pentagons-race-against-deepfakes>

디지털과 모바일

BBC는 STT(speech-to-text) 기술을 통해 뉴스 스크립트를 분석하고, 스크립트의 일부분을 선택하여 바로 온라인 기사용 클립을 만들 수 있는 ‘Online Content Tool’을 선보였다. KBS 미디어기술연구소도 AI 기반 영상 분석을 통해 특정 인물이 등장하는 클립을 자동 생성할 수 있는 ‘알파클립’ 프로토타입 시스템을 개발한 바 있다. 온라인 디지털 영상 소비가 늘어남에 따라 이러한 클립 생성 시스템의 필요성은 갈수록 증가할 것으로 보인다.

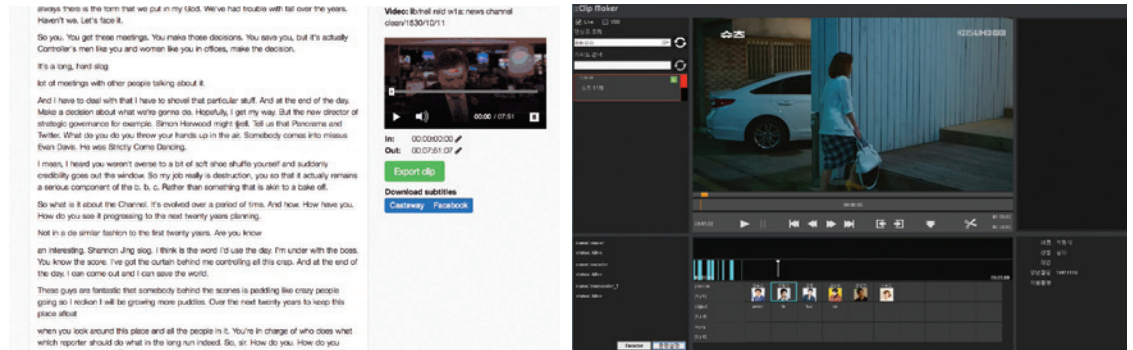


그림 5. 클립 생성 시스템. (좌)BBC의 Online Content Tool, (우)KBS 알파클립

KBS 미디어기술연구소는 또한 모바일 디바이스에서 화면을 돌릴 필요가 없는 ‘세로 영상’이 대세임에 착안하여, 아이돌 무대 영상의 세로 ‘직캠’ 영상 등을 빠르게 생성할 수 있는 ‘VERTIGO’ 시스템을 개발하였다. VERTIGO 시스템은 아이돌 무대 영상을 입력받아서 AI를 이용해 얼굴을 검출하고 추적하며 멤버별로 적절한 카메라 프레임 을 동적으로 생성한다. 사실 아이돌 무대에서 멤버별 ‘직캠’ 영상을 직접 촬영하는 것은 어지간한 ‘팬심’이 아니고서는 쉽지 않다. 일단 원하는 아이돌 멤버 숫자만큼 카메라 설치 및 촬영 스태프 배치가 필요하고, 촬영자가 군무를 어느 정도 파악하고 있지 않다면 멤버를 프레임에 담아 따라가다가 놓치기 십상이다. 따라서 VERTIGO 시스템으로 전체 멤버를 촬영한 후 오프라인에서 멤버별로 분할하게 되면 많은 직캠 촬영 스태프와 장비를 들이지 않으면서도 모든 멤버별 영상을 얻을 수 있는 장점이 있다. 개발 중인 이 시스템은 음악 프로그램 등의 세로 영상 제작에 적극 활용 예정이다.

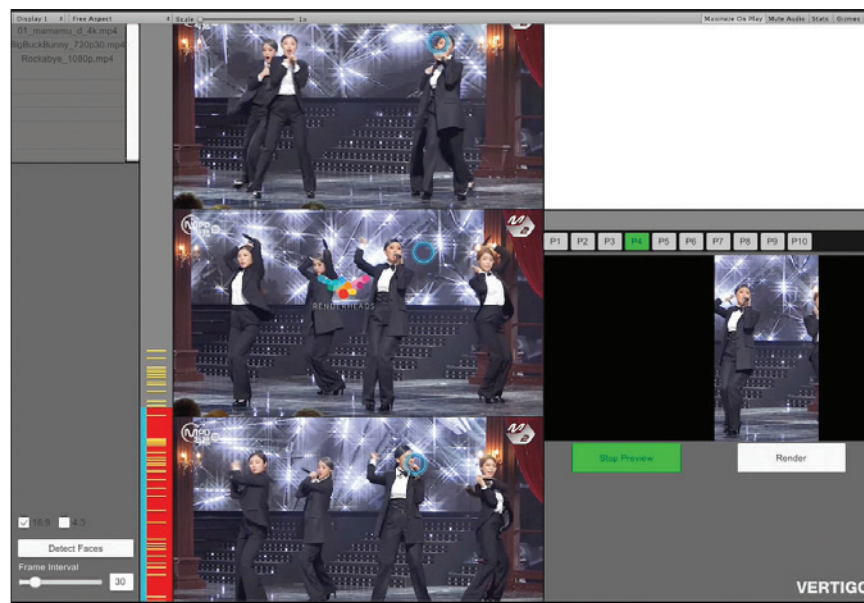


그림 6. 특허 출원 중인 VERTIGO 시스템의 UI

이상으로 뉴스 프레젠테이션, 품질관리, 메타데이터와 검색, 자동 제작, 미디어 포렌식, 광고, 디지털과 모바일 등 대표적인 AI의 방송 미디어 응용에 대해 살펴보았다. 어떤 응용은 지금도 AI가 적용 및 활용되고 있고, 어떤 응용은 아직 좀 더 시간이 필요한 것들도 있다. 이 중에서 AI 도입으로 가장 효율성이 극대화될 것으로 예상되는 대표적인 응용은 메타데이터 생성일 것이다. 이제 메타데이터 생성과 같은 응용을 위해, AI 기술을 도입하는 방식에 대해 방송사의 입장에서 논의해 보고자 한다.

지금은 클라우드가 대세다. 적은 도입 비용과 총소유 비용(TCO), 유연성과 효율성 등 클라우드의 장점은 널리 알려져 있다. 그러다 보니 많은 기업이 서버 등의 인프라(즉 IAAS)뿐 아니라 AI와 같은 고급 기능의 사용에서도 클라우드 도입을 검토하고 있다. 실제로 별도의 연구 투자 없이도 금세 결과를 보여주는 클라우드 AI는 대단히 매력적이다.

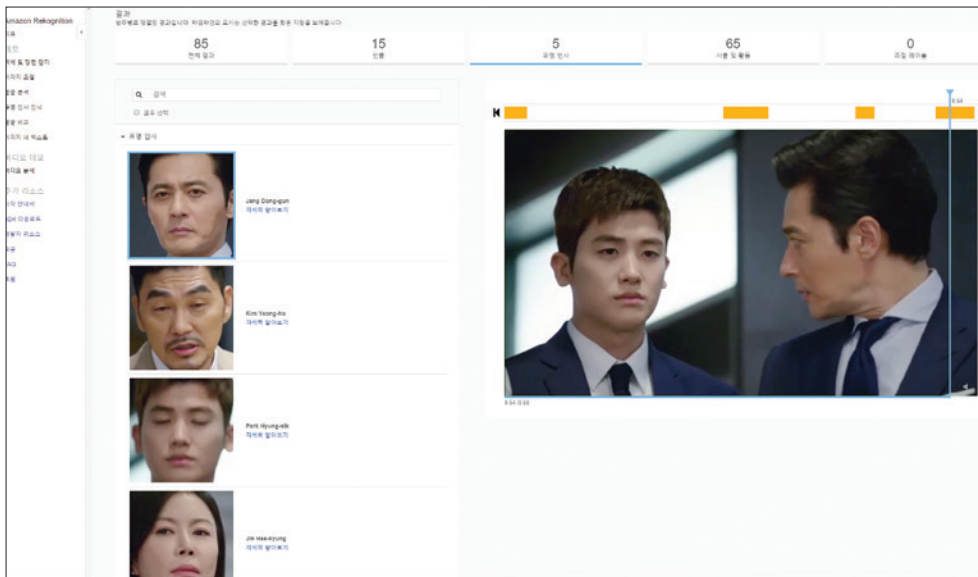


그림 7. AWS Rekognition을 통해 KBS 드라마 '슈츠'의 일부분을 분석한 결과

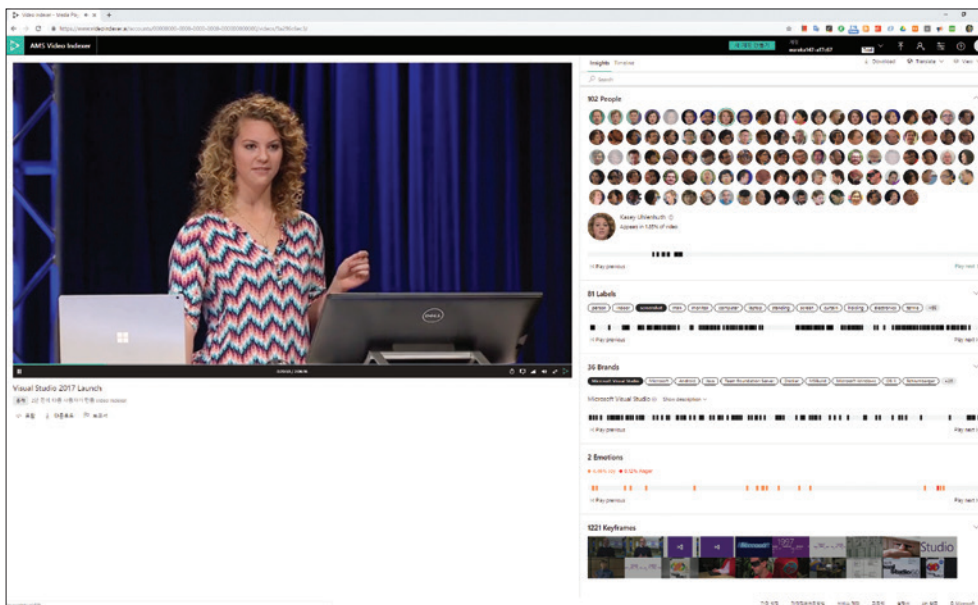


그림 8. Microsoft Video Indexer의 영상 분석 결과

[그림 7, 8]에서 보듯이 많은 연구개발 투자가 이루어진 클라우드 AI는 꽤 정확한 결과물을 내놓는다. 그러나 이는 유명인(celebrities) 중에서도 톱클래스 출연자에 한정된 결과라 할 수 있으며 실제로 방송에 등장하는 인물들은 생각보다 매우 다양하여 톱클래스와는 거리가 먼 경우가 대다수를 차지한다. 국내의 다양한 방송 미디어 등장 인물들을 모두 학습하여 아카이브 영상에 메타데이터를 부착하는 것을 목표라고 할 때 클라우드에서 제공하는 AI는 이에 적합하다고 하기는 어렵다. 국내의 수많은 출연자에 대한 검출도 어렵지만 비용 문제도 발생한다. KBS의 경우 약 64만 시간 분량의 디지털 영상 아카이브가 존재한다. 클라우드 AI API를 통해 이를 모두 분석하려고 하면 엄청난 비용이 발생할 수밖에 없다. 방송사 자체(in house) 솔루션이 필요한 이유이다.

AI도 결국 분류기(classifier) 개발로 귀결된다. 분류기에서 가장 중요한 것은 원 데이터로부터 특징들(features)을 뽑아내어 분류 가능한 특징 공간(feature space)에 놓는 것이다. 딥러닝 이전의 머신러닝에서는 연구자가 특징 추출을 어떻게 하는지에 해당하는 알고리즘이 성능에 큰 영향을 미쳤다. 그러나 지금의 AI는 방대한 학습 데이터로부터 특징을 네트워크가 스스로 찾아내는 딥러닝 기반이다. 딥러닝이 도입되면서 세부 알고리즘 간의 성능 차이는 크지 않게 되었다. 즉 지금 AI 개발에서 가장 중요한 것은 알고리즘이 아닌 데이터셋이라는 의견이 지배적이다.

전 세계적으로 AI 개발이 큰 화두인 지금, 방송사들은 AI 데이터셋의 기반이 되는 콘텐츠를 보유하고 있음에 주목할 필요가 있다. 물론 대부분 레이블링되지 않은 데이터(unlabeled data)이지만 학습용 초기 데이터는 그리 크지 않은 비용으로 레이블링할 수 있을 것으로 보인다. 일단 어느 정도 수준의 AI가 구축되면, 그 후에는 대부분은 AI 분석을 통해서, 그리고 일부 오류는 사람의 검수를 통해 보완하는 식으로 추가적인 데이터셋을 계속 확보할 수 있을 것이다. 사실 메타데이터 요원이 콘텐츠에 메타데이터를 부착하는 것 자체가 AI 학습 데이터셋의 확보이다. 즉 메타데이터 부착(레이블링)과 AI 개선 사이의 지속적인 선순환을 기대할 수 있는 것이다.

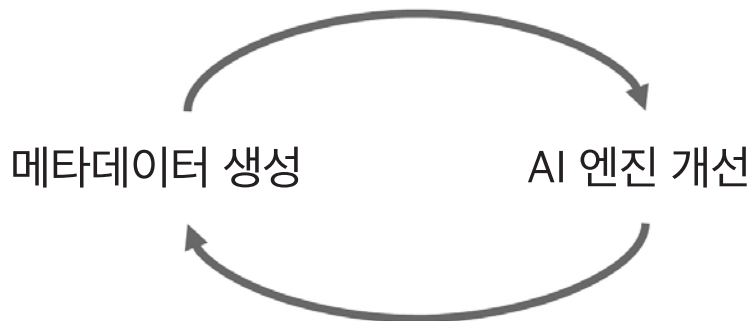


그림 9. 메타데이터 생성과 AI 엔진 개선의 지속가능한 선순환 구조

추가로 방송사는 영상뿐 아니라 자막 등 부가적인 정보도 보유하고 있다. 이를 이용하면 한국어 음성인식 AI 엔진 개발 등도 가능하다. 즉 아직은 방대한 콘텐츠를 보유한 방송사가 적어도 AI 개발에서는 큰소리를 낼 수 있는 상황이라는 뜻이다. 따라서 국내 방송사에서 미디어 콘텐츠 분석을 포함한 AI 기능을 도입할 예정이라면 맹목적인 클라우드 도입보다는 인하우스 개발을 우선적으로 검토할 것을 권하고 싶다.

이상으로 이슈가 되고 있는 방송 미디어를 위한 AI 기술의 개요 및 현황과 응용, 그리고 도입 방안에 대해 2부에 걸쳐 살펴보았다. 시간을 내어 읽어 주신 현장의 독자들에게 본 내용이 조금이나마 도움이 되기를 바라며 글을 마무리하고자 한다. 📖