

데이터 분석 결론

원본 CSV, XML, JSON 세 가지를 비교해 본 결과,

1. CSV와 XML은 같은 환자의 데이터, JSON은 다른 환자의 데이터입니다.
2. XML 데이터 상, 검사 시 Vivid S60 장비를 사용한 것으로 파악되며 Vivid S60은 GE Healthcare 사의 초음파 기기입니다. JSON은 Siemens 사의 장비를 사용한 것을 이미 알고 있으므로 XML/JSON 두 자료의 차이는 피 검사자와 검사장비이며, 이 차이로 인해 측정 파라미터와 측정 값, 데이터 구조 등이 다를 것이라고 판단하였습니다.
3. 데이터 분석을 보다 효율적으로 하기 위해 XML과 JSON 파일에서 작성한 코드로 측정 파라미터와 측정 값을 따로 추출하였으며, 보고서에서 설명드렸던 대로 기기 종류와 상관없이 필수적으로 측정하는 값들이 있는 것을 확인하였습니다. 이 값들이 어떤 검사결과값을 나타내는지 찾아서 간단하게 정리해보았으며, 결과적으로 심장 초음파와 관련된 값이라는 것을 알게 되었습니다.
4. 심장/동맥/폐를 검사하여 낸 계산값이 대부분 중요 데이터로 들어와 있는 것을 확인할 수 있었는데, 해당 부위의 두께, 크기, 해당 부위를 지나가는 동맥 또는 정맥혈의 혈류 속도, 크기 비율, 부위 간의 압력차 등 심장초음파로 측정할 수 있는 다양한 값이 측정계산된 것을 파악할 수 있었습니다. 해당 부위 자체를 검사하는 것은 주로 MMode나 2D Mode, 혈류와 관련된 검사는 대부분 도플러 모드로 수행하는 것을 알 수 있었습니다.
5. MMode로 측정할 수 있는 값들 중 2D Mode로도 측정 가능한 값이 있고, 그렇지 않은 값이 있는 것 같다고 추측하였습니다. 2D Mode로 측정 가능한 값 역시 MMode로 측정할 수 없는 값이 있는 것 같다고 추측하였습니다.
6. CSV, XML 파일만을 비교했을 때는 CSV 파일 기준 NULL 값이 들어와 있는 검사항목은 XML 파일에 존재하지 않았으며 CSV 파일 기준 NULL값이 아닌 항목은 XML 파일에 전부 들어와 있는 것을 알 수 있었습니다. CSV 파일의 데이터가 HL7 포맷의 메시지로 보내는 기준 항목일 것일 것이라고 추측해보았습니다. 또한 CSV 파일의 데이터가 HL7 포맷의 메시지로 보내질 때 Result No. 항목의 값도 어떠한 기준점이 되는 것 같다고 추측했지만, 정확한 규칙은 찾지 못하였습니다. 마찬가지로 JSON 파일이 HL7 메시지로 보내질 때 어떤 형식의 CSV 파일로 전송되는 지 알게 된다면 더욱 상세한 분석이 가능할 것이라고 생각합니다.
7. 세 파일을 모두 비교했을 때 각 파일마다 어떤 파일에도 겹치지 않는 항목 또한 존재했는데, 이러한 항목의 경우 데이터 구조상 아직 모든 파라미터의 의미를 알 수 없으므로 조금 더 많은 정보가 있어야 왜 겹치지 않는지에 대해서 추측해볼 수 있을 것 같습니다.
8. JSON 데이터의 Redirect 부분 또한 언제 어떻게 Redirect가 되는 지에 대해서는 자세히 알 수 없었습니다. 그룹화된 특정 값들을 Redirect 부분에서 다시 분류하고 있는 것이 확인되어 Redirect 부분을 코드 내에서 분리 파싱하고 있기는 하지만 이 부분에 대해서는 일관된 규칙이 존재한다고 보기에는 어렵다는 판단이었고, 이 부분 역시 조금 더 많은 정보가 있어야 추측이 가능할 것 같습니다.