**Development of Mass Spectrometry Data Plotting Software Using Python**

20121392 Hwayong, Jang

**1.Project Title:**

파이썬을 이용한 질량분석 데이터 플로팅 소프트웨어의 개발

**2.Project Background:**

화학 분야에서 자주 쓰이는 분석 방법 중 하나인 ‘Mass Spectrmoetry(질량분석)’는 화합물의 분자 구조를 분석하는데 사용됩니다.  
  
기존의 방법으로 질량분석기를 통해 얻은 질량분석 데이터를 분석하려면 제조사가 판매하는 질량분석 소프트웨어 <Xcalibur>를 사용해야합니다. 하지만 <Xcalibur>는 판매가격이 높을 뿐 만 아니라 추가  
적인 비용을 지불해 정기적 업그레이드를 해주어야 한다는 단점이 있습니다. 또한 인공지능을 이용한 분석 등 분석자가 시도하고자 하는 데이터 처리를 위해 프로그램을 re-modeling하는 것이 불가능  
합니다.  
  
이러한 이유로 질량분석학계에서는 제조사가 제공하는 질량분석용 소프트웨어 대신 분석자가 무료로 이용 가능하도록 질량분석 소프트웨어를 개발중인 단계입니다.

**3.Problem Statement:**

\* 분석은 대부분 인터넷이 연결되어 있지 않은 offline상태의 컴퓨터를 통해 분석되므로, 데이터를 플로팅하는데 있어 네트워크를 이용하는 plotly등의 라이브러리를 사용할 수 없음.

\* 질량분석 데이터는 텍스트로 200mb정도의 큰 데이터이므로 빠른 연산을 위해 최적화된 모듈을 사용해야한다.

**4.Objectives:**

**1. 데이터 추출**

\* mzXML 파일을 처리하기 위해 질량분석을 위한 파이썬 오픈소스 라이브러리 pyOpenMS를 이용

\* mzXML형식의 질량분석 데이터파일에서 BPS(Base Peak Chromatorgam)와 MS(Mass Spectrum)를 플로팅하기 위한 데이터를 추출

\* 이때 필요한 데이터는 다음과 같다.

Base Peak Chromatogram:  
X축 = retention time  
Y축 = base peak intensity

Mass Spectrum:

X축 = m/z

Y축 = intensity

\* 추출한 데이터를 NumPy와 pandas를 이용하여 저장

**2. 데이터 플로팅**

\* matploblib을 이용, BPS와 MS를 플로팅한다.

**5.Expected Benefits:**

1. 비싼 가격의 기존 분석소프트웨어 대신 분석자 누구나 무료로 이용 가능한 소프트웨어를 개발 할 수 있다.

2. 기존의 분석소프트웨어는 최신 연구 동향인 인공지능을 활용한 데이터분석을 위해 소프트웨어  
를 re-modeling할 수 없지만, 직접 개발한 이번 프로젝트 소프트웨어의 경우 필요한 분석을 위해 소프트웨어를 re-modeling할 수 있다.