

반응 정보를 포함한 오픈 데이터의 Crawling에
기반한 스마트 반응경로 생성 및 반응위험성
분석 시스템

정준수, 박소현, 신동일[†]

명지대학교

(dongil@mju.ac.kr[†])

오늘날 웹상에는 매우 많은 데이터가 축적되어 있다. 그중 오픈된 화합물 데이터, 반응 데이터, 안전 데이터 등을 포함하여 유용한 데이터가 많다. 연구자들은 실험 시작 전 또는 공정 설계 전 이러한 정보들을 찾아보는 것이 필수적이다. 하지만 소스들이 웹상에 광범위하게 퍼져 있어 검색이 어렵고, 시간도 오래 걸린다. 제안된 시스템은 연구자들이 반응 정보를 검색하고 위험성을 분석하는 것을 지원한다. 시스템에 사용되는 DB는 PubChem과 같은 오픈된 페이지에서 찾을 수 있는 화합물, 반응 및 안전 등에 대한 오픈 데이터로 구성된다. DB 구축은 Python 기반의 Web Scraping 및 Web Crawling을 활용해 자동적으로 진행되었다. 반응경로 생성을 위해서는 depth-first search 알고리즘과 pruning 알고리즘이 조합하여 사용했다. 그 결과 탐색이 진행되면서 가능성이 없다고 판단되는 branch는 사전에 제외되고 작동하게 된다. 위험성 분석은 열역학적 데이터를 가지고 반응 시 생길 잠재적인 위험성 확인과 물질 자체가 갖는 유해성 데이터를 통해 진행되었다. 이를 통해 생성된 결과는 시작물질부터 목표물질까지 진행되는 다양한 경로들을 간단하게 표현된 노드로 보여준다. 각 경로 상 예상되는 위험성은 별도의 탭에서 나타난다. 생성된 결과를 참조하여 연구자들은 적은 시간으로 경제성이 고려된 다양한 반응경로를 찾고, 아울러 실험 사고를 예방할 수 있을 것으로 기대된다.