

오픈데이터 자동 수집과 데이터 마이닝 기반의 반응경로 자동생성 시스템

정준수, 신동일[†]

명지대학교

(dongil@mju.ac.kr[†])

실험실, 파일럿 플랜트, 공정 등에서 목표물질을 설정하고 반응경로를 디자인하기 위해서는 관련 정보를 찾아보는 것이 필수적이다. 이러한 정보는 인터넷과 문헌들을 통해 찾아볼 수 있지만, 상업용 데이터를 활용하지 못하면 정보 수집이 매우 어렵다. 또한 반응물, 생성물, 촉매, 수율, 시간, 온도, 독성, 위험성 등 다양한 분류의 정보를 각각 탐색해야 한다. 때문에 목표물질에 대한 안전한 반응경로를 디자인하기가 쉽지 않다. 본 연구에서는 인터넷의 관련 오픈데이터들을 Web Crawling을 통해 수집하고, 데이터 마이닝을 진행해 지식그래프로 만들었다. 이를 통해 목표물질을 합성할 수 있는 다양한 반응경로들을 자동으로 탐색하고 생성한다. 지식그래프 구축은 그래프 데이터베이스인 AllegroGraph를 사용하여 진행하였다. 반응경로 생성 시 각 반응 및 반응물과 생성물에 대한 독성, 유해성, 에너지변화 등을 계산하여 각각의 경로들을 평가한다. 제안된 시스템은 반응경로 그래프와 각 경로들에 대한 상세 정보를 나타내는 두 가지 결과를 보여준다. 이를 통해 연구자들이 반응경로 디자인 시 제안된 시스템을 이용해 관련 반응경로들을 생성하면 반응경로 선택에 투자해야만 하는 시간을 대폭 줄이는 효율적인 시스템이 될 것으로 기대된다.