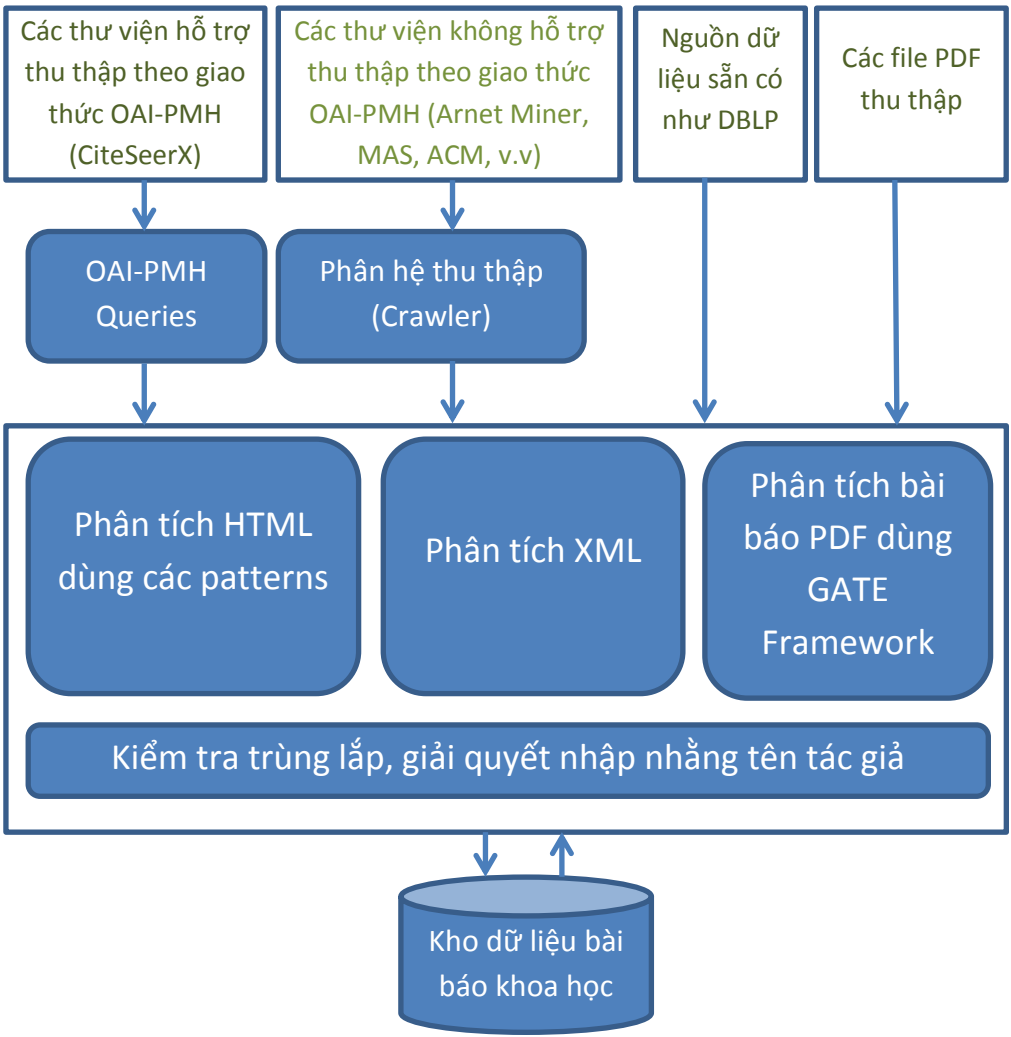
Các thư viện hỗ trợ thu thập theo giao thức OAI-PMH (CiteSeerX)
Các thư viện không hỗ trợ thu thập theo giao thức OAI-PMH (Arnet Miner, MAS, ACM, v.v)
Nguồn dữ liệu sẵn có như DBLP
Các file PDF thu thập
OAI-PMH Queries
Phân hệ thu thập (Crawler)
Phân tích HTML dùng các patterns
Phân tích XML
Phân tích bài báo PDF dùng GATE Framework
Kiểm tra trùng lắp, giải quyết nhập nhằng tên tác giả
Kho dữ liệu bài báo khoa học