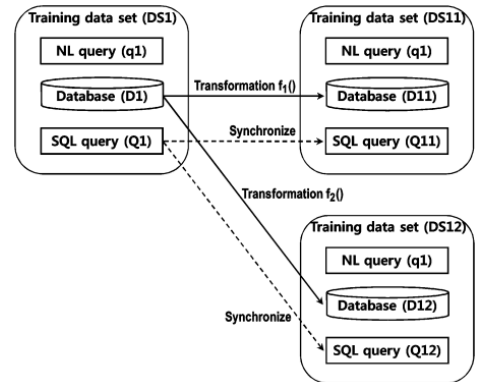


자연어 처리 시스템의 학습 데이터셋 증강 방법

POSTECH 전자전기공학과 송호진

기술의 개요

- SQL-to-Text 모델을 학습시키기 위한 다양한 데이터를 수집 혹은 생성하는 것에 대한 어려움이 여전히 존재
- 수작업으로 생성한 템플릿을 요하지 않으면서, 자연어 학습 데이터의 추가가 없더라도 다양한 학습 데이터를 증강시킬 수 있게 해주는 학습 데이터 증강 방법



[그림. 학습 데이터셋 증강 과정]

기술의 특징 및 장점

데이터 다양성 확대

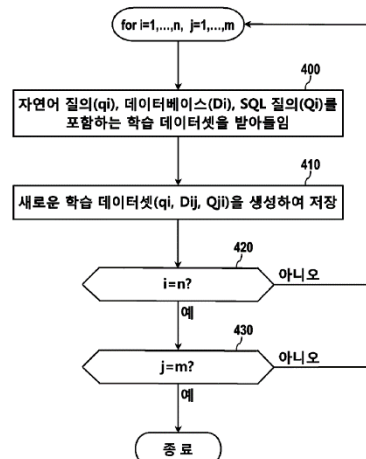
수작업으로 템플릿을 생성하지 않더라도, 데이터베이스 스키마 변형 연산을 통해 학습 데이터를 자동으로 생성

다양한 학습 시나리오에 적용 가능

데이터베이스 스키마를 처리할 수 있어, 특정 도메인에 한정되지 않고 다양한 학습 시나리오에 적용 가능

기술 경쟁력

- 데이터베이스 스키마의 변형을 통해 다양한 학습 데이터를 자동으로 생성함으로써, **기존의 수작업 방식보다 훨씬 효율적**
- 다양한 스키마를 통해 생성된 데이터는 신경망 모델이 더 넓은 범위의 질의를 처리할 수 있도록 학습시켜, **시스템의 일반화 능력을 크게 향상**



[그림. 학습 데이터셋 증강 동작 과정]

기술 성숙도 (TRL*)

*TRL : Technology Readiness Level

| | | | | | | | | |
|------|---|----|---|-----|---|-----|---|-----|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| 기초연구 | | 실험 | | 시작품 | | 실용화 | | 사업화 |

- 연구실/실험실 규모의 환경에서 기본 성능이 검증될 수 있는 단계
- 개발하려는 시스템/부품의 기본 설계도면을 확보하는 단계
- 모델링/설계 기술 확보

시장동향

- **시장전망 : 글로벌 데이터 품질관리 시장 연평균 18.7% 성장 기대**
- ✓ 글로벌 데이터 품질관리 시장은 2021년 11억 4,000만 달러의 시장 규모를 달성하였으며, 연평균 18.7% 성장하여 2027년 33억 달러에 이를 것으로 전망됨

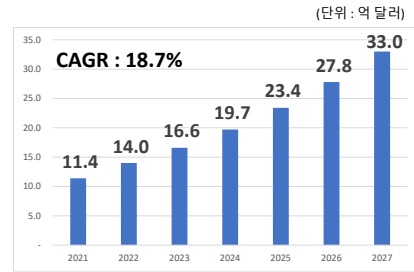


그림. 글로벌 데이터 품질관리 시장
자료: Data Quality Tools - Global Strategic Business Report_ Research And Markets_('24.01)

활용 분야 및 적용 제품

데이터
관리



- 데이터관리 소프트웨어
- 기업 데이터관리 및 분석 시스템
- 자연어 질의 기능 제품

교육 및
학습도구



- 교육용 소프트웨어
- SQL 학습 지원

데이터관리 및 교육분야에 적용이 가능
데이터관리 및 SQL 학습 소프트웨어에 사용 가능

지식재산권 현황

| 발명의 명칭 | 출원번호 | 권리현황 |
|---------------------------|-----------------|------|
| 자연어 처리 시스템의 학습 데이터셋 증강 방법 | 10-2023-0102382 | 심사중 |

문의처

기술
이전

기술거래기관 (주)피앤아이비
PHONE 070-8233-5678
E-MAIL pnib@pnibiz.com