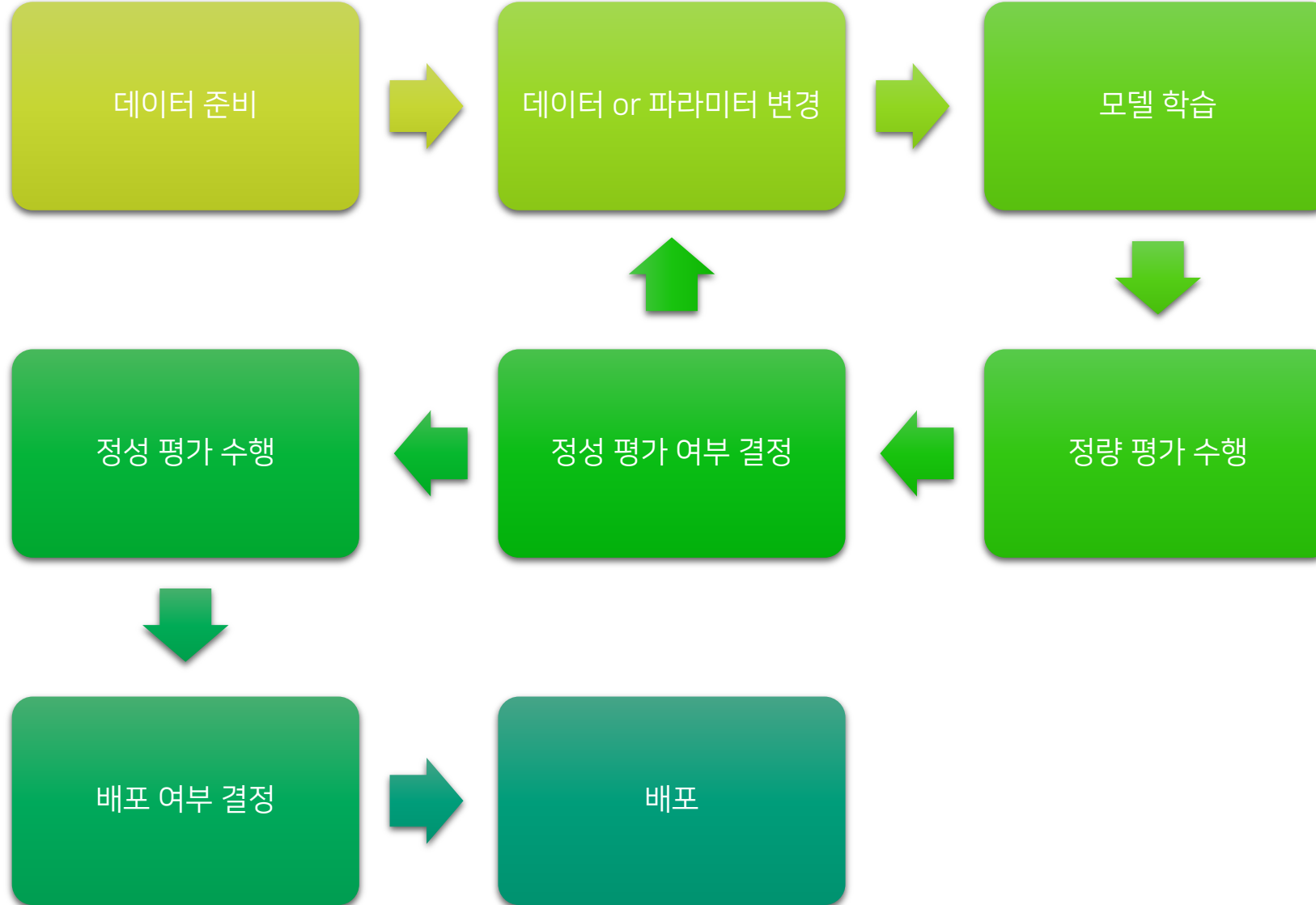


TIP: How to Evaluate

Ki Hyun Kim

nlp.with.deep.learning@gmail.com

배포까지의 과정



How to measure BLEU for each Language

- 공통 작업
 - BPE 등의 tokenization을 모두 detokenization
- 한국어
 - Mecab등을 통해 접사 및 특수 문자(e.g. 마침표, 괄호 등)를 분리 후, BLEU 측정
- 영어 (라틴어 기반의 언어)
 - Moses등을 통해 특수 문자를 분리 후, BLEU 측정
- (띄어쓰기가 없는) 중국어, 일본어
 - Mecab, Jieba 등을 통해 segmentation을 수행 후, BLEU 측정

How to get intrinsic evaluations

- 잘 선정된 test set을 각 경쟁사(?) 번역기에 넣어 평가
 - 절대 평가
 - 실제 번역 품질을 여러 단계로 나누어 평가
 - 일관적 평가 품질 유지가 관건
 - 상대 평가
 - 각 벤치마크 대상들을 모아 놓고, 순위를 매겨 평가
 - 객관적 평가 품질 유지가 관건
- 정성 평가를 수행 할 때 유의할 점:
 - Blind test
 - 평가의 용이성 확보
 - 평가 리소스 확보