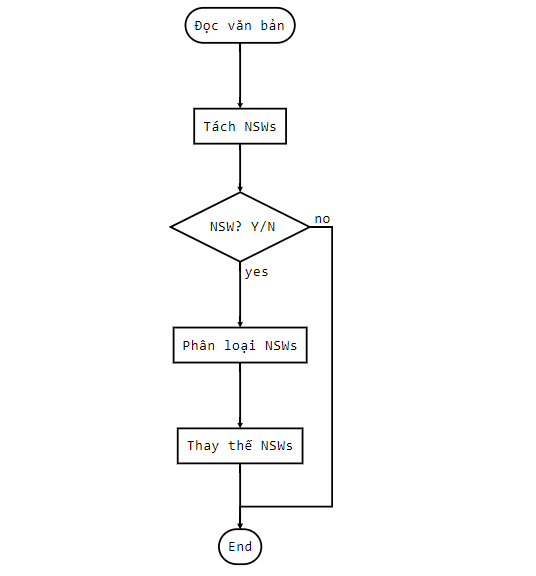
**Normalize-text for TTS**

**1. Active Flow**



**2. Requirements**

- num2words

- beautifulsoup4

- lxml

- pronouncing

- tqdm

- unicodedata

**3. Function**

- config: chứa các dict và luật để phân loại và đọc các từ dạng NSW

- create\_dict: tạo thư viện đọc từ tiếng anh

- en2vi: hàm xây dựng âm tiết với các từ tiếng anh(hiện đang sử dụng en2vi\_old)

- spit\_token: tách riêng các NSW và các từ đọc được

- expand\_NSWs: hàm đọc các loại NSW đã phân loại

- classify\_token: hàm phân loại các NSW

- replace: thay thế các NSW sang dạng đọc được

- normalize-text: main

**4. Maintain**

- config: thêm dict, rule

- split\_token: tách các token là NSW và các từ đọc được bằng regrex

- expand\_NSWs: diễn giải các NSWs theo từng dạng được phân loại

- classify\_token: phân loại các NSW vào 3 group và 21 category

- folder dicts: chứa các thư viện sử dụng khi diễn giải NSWs và được đặt tên theo các dạng NSW từ file normalize\_text.doc