**BÁO CÁO DỮ LIỆU**

**TỪ ĐIỂN TRỰC TUYẾN**

**Nhóm: 13**

*Hà Nội ngày 15, tháng 03, năm 2019*

Nhóm thực hiện xây dựng, tinh chỉnh lại dữ liệu nội dung, âm thanh từ ngữ từ nguồn không tính phí:

* **Dữ liệu nội dung từ ngữ:**
  + Sử dụng bộ từ ngữ tại:
    - <http://www.informatik.uni-leipzig.de/~duc/Dict/> *(hiện tại không thể truy cập được)* tải về ở dạng file text.
  + Thực hiện tách bộ từ thành từ thông thường và từ chuyên ngành (*hiện tại đang sử dụng bộ từ thông thường*).
  + Thực hiện xây dựng lại bộ từ thành tệp tin có cấu trúc ***.dict*** được mã hoá và giải mã theo tệp tin ***thu\_vien.php***.
* **Dữ liệu âm thanh từ ngữ:**
  + Sử dụng ngôn ngữ **Python**:
    - Sử dụng thư viện **gTTS** tại <https://pypi.org/project/gTTS/> để thực hiện lấy tệp âm thanh tương ứng với các từ có trong dữ liệu.
    - Thực hiện lưu trữ các tệp âm thanh trên Github repository để sử dụng.