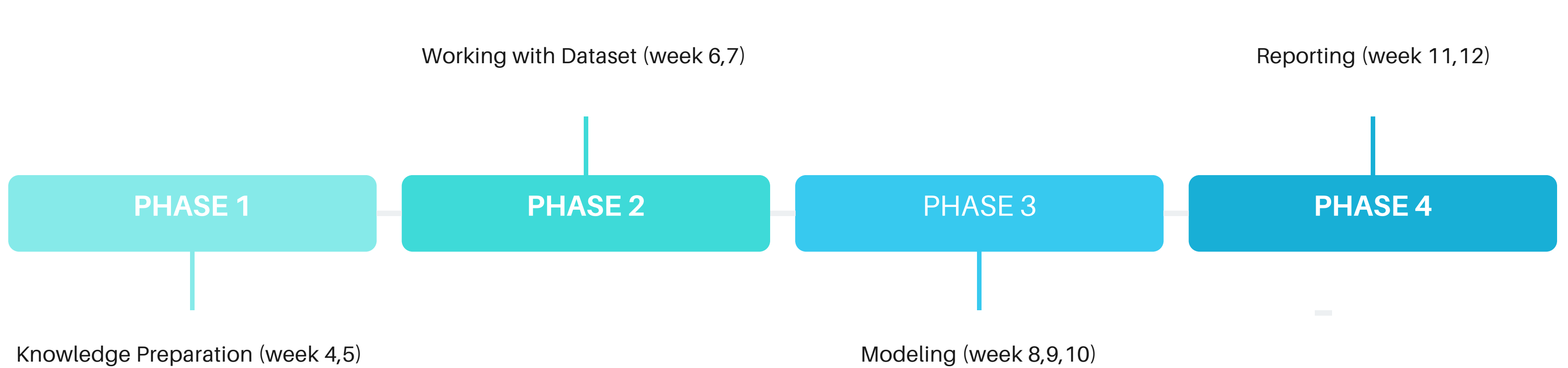


***Week 4* REPORT**

1. TIMELINE



PHASE 1

1. Học Git (due 2/6):

- Sử dụng công cụ Git, học thao tác các lệnh như git.push, git.commit, git add... để có thể dễ dàng làm việc trên Github

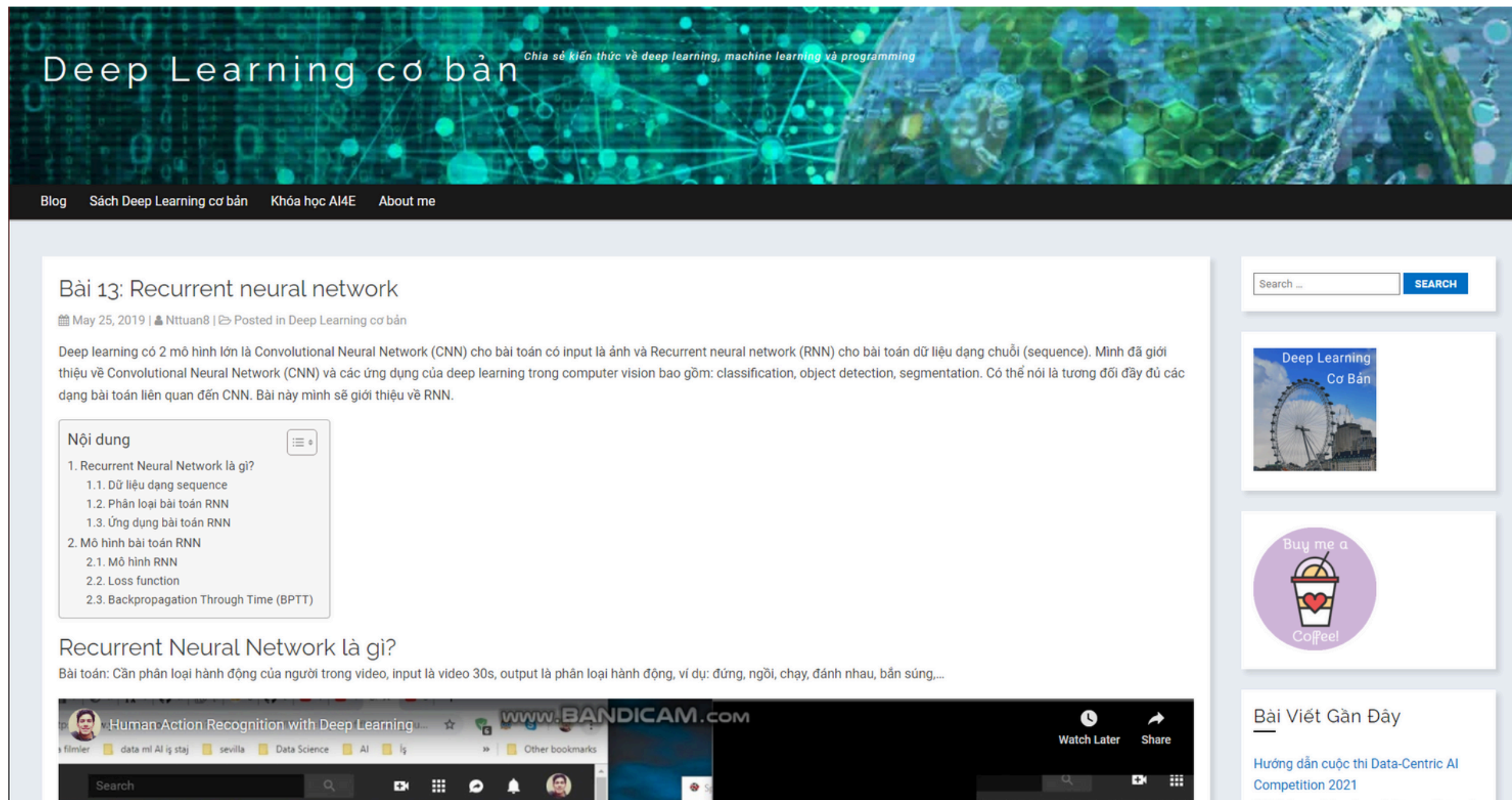
2. Học lý thuyết (due 9/6):

- Nghiên cứu lý thuyết về [RNN](#) và những khóa học liên quan như linear vs logistic để củng cố

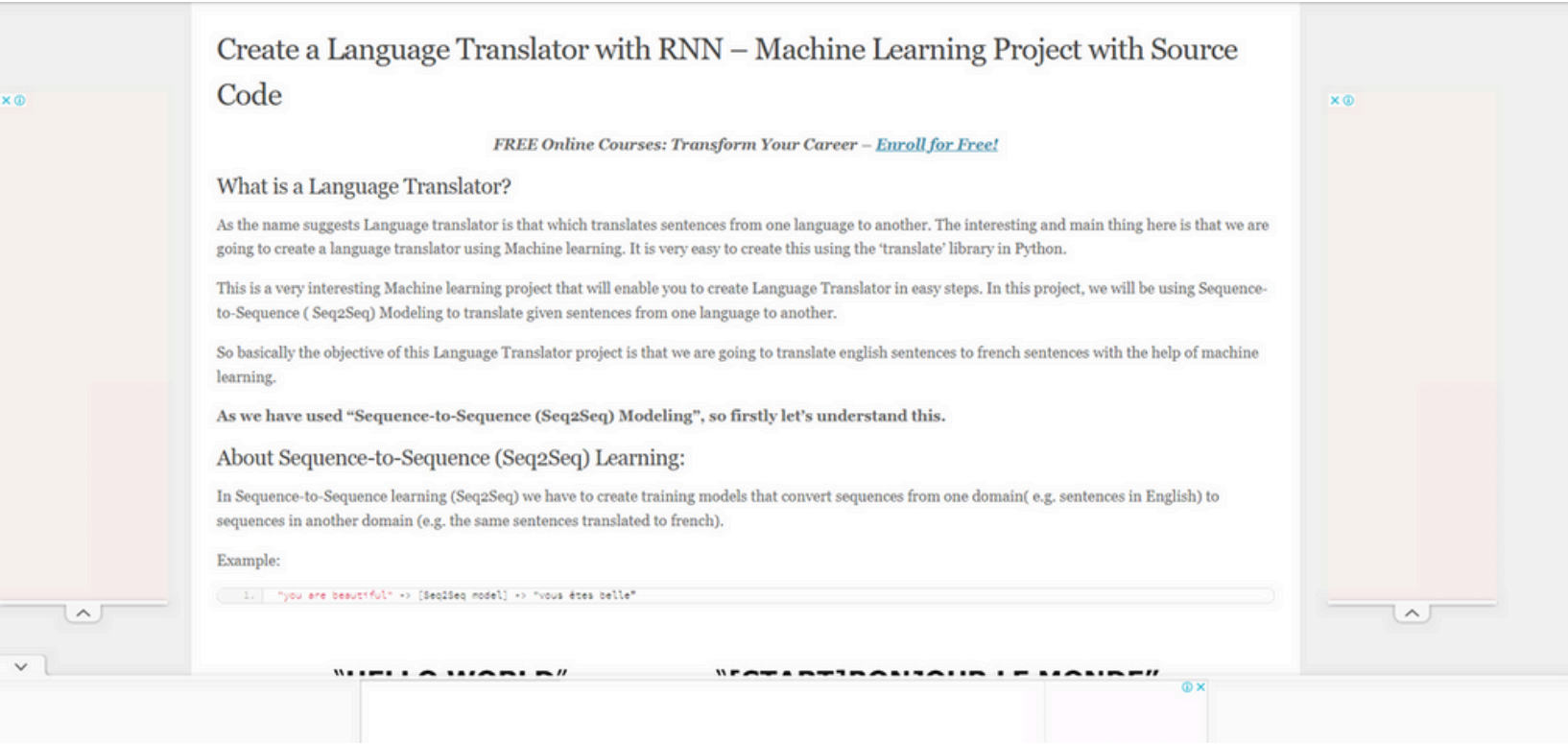
3. Hands-on Project (due 9/6):

- Tìm được [guide](#) có đầy đủ thông tin cần thiết để tham khảo code với dataset

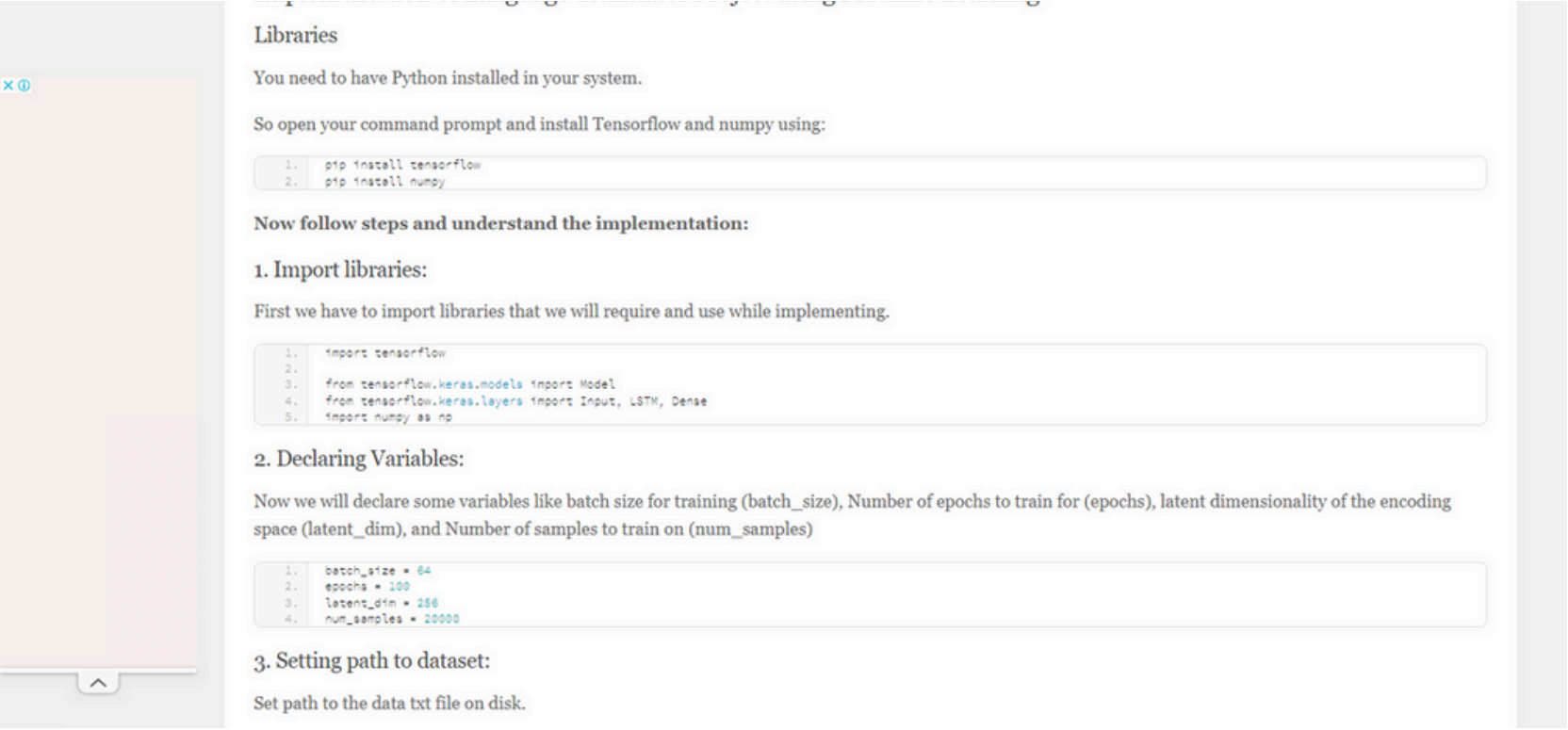
2. GUIDE



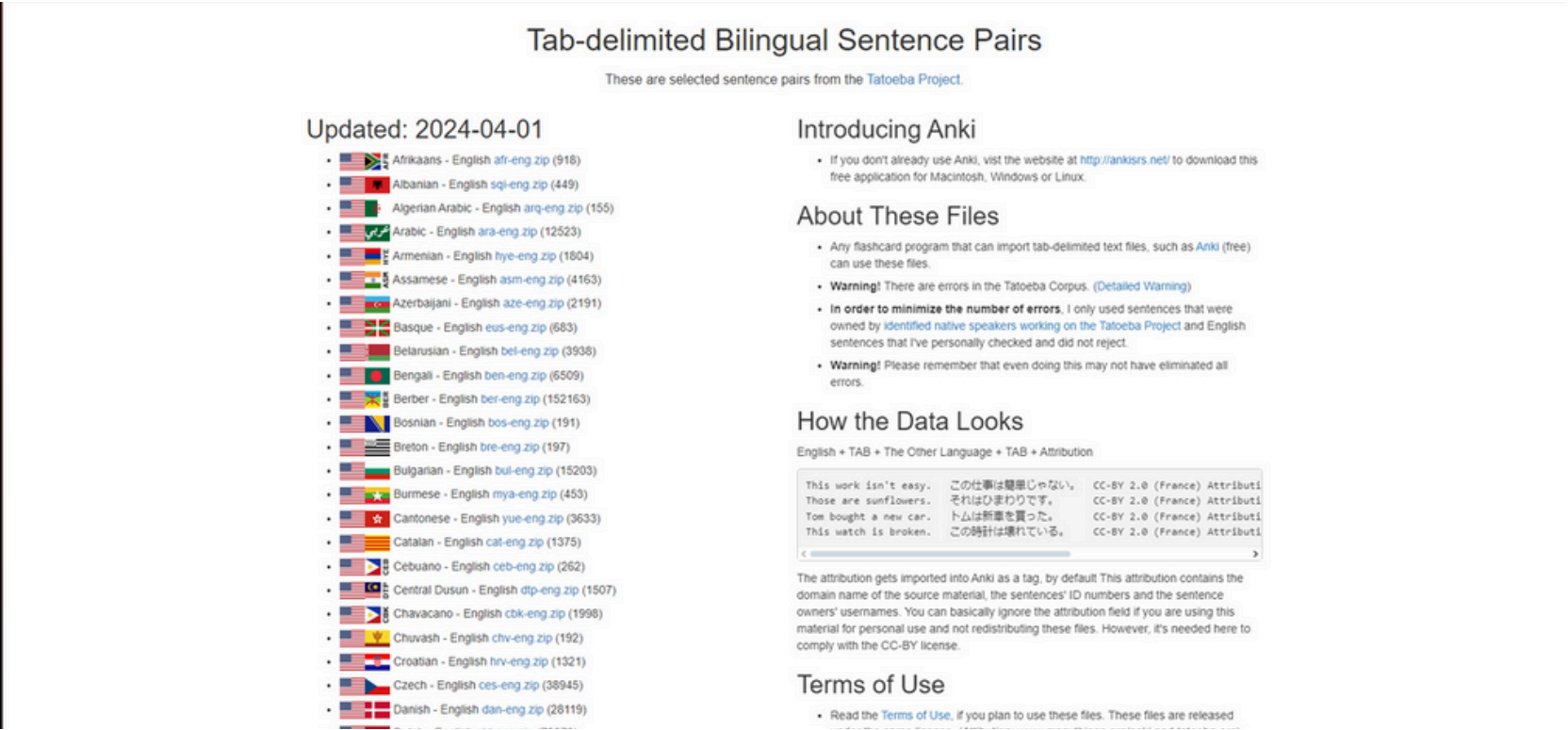
2. GUIDE



Cung cấp kiến thức cơ bản về lí thuyết




Vừa có cả code trực quan



Và cả dataset để thực hành

3. DATASET



English Vietnamese

Q

☀

Contribute

Publications

Corpora

Resources for English (en) - Vietnamese (vi) - (20 found)

Corpus	sentences ^	en tokens	vi tokens	sample	bilingual	monolingual
CCMatrix v1	50,092,444	790,137,331	1,082,123,673	👁	moses <div>↓ 🔗</div>	txt en <div>↓ 🔗</div>
NLLB v1	50,092,444	790,153,460	1,082,140,912	👁	moses <div>↓ 🔗</div>	txt en <div>↓ 🔗</div>
CCAligned v1	12,394,417	131,721,034	178,824,192	👁	moses <div>↓ 🔗</div>	txt en <div>↓ 🔗</div>
OpenSubtitles v2018	3,505,276	21,218,934	25,651,558	👁	moses <div>↓ 🔗</div>	txt en <div>↓ 🔗</div>
WikiMatrix v1	1,073,752	22,545,174	30,787,586	👁	moses <div>↓ 🔗</div>	txt en <div>↓ 🔗</div>
wikimedia v20230407	669,743	15,814,411	20,577,828	👁	moses <div>↓ 🔗</div>	txt en <div>↓ 🔗</div>
QED v2.0a	338,024	4,239,747	5,434,287	👁	moses <div>↓ 🔗</div>	txt en <div>↓ 🔗</div>
TED2020 v1	326,417	5,536,190	7,073,857	👁	moses <div>↓ 🔗</div>	txt en <div>↓ 🔗</div>
NeuLab-TedTalks v1	184,973	3,340,654	4,166,860	👁	moses <div>↓ 🔗</div>	txt en <div>↓ 🔗</div>
StanfordNLP-NMT v1.0	133,167	3,019,542	3,310,135	👁	moses <div>↓ 🔗</div>	txt en <div>↓ 🔗</div>
ELRC-wikipedia_health v1	126,413	4,274	91,679	👁	moses <div>↓ 🔗</div>	txt en <div>↓ 🔗</div>
bible-uedin v1	124,390	3,100,862	3,425,151	👁	moses <div>↓ 🔗</div>	txt en <div>↓ 🔗</div>
Wikipedia v1.0	58,116	488,139	487,761	👁	moses <div>↓ 🔗</div>	txt en <div>↓ 🔗</div>
KDE4 v2	42,782	215,523	287,089	👁	moses <div>↓ 🔗</div>	txt en <div>↓ 🔗</div>
Tatoeba v2023-04-12	6,855	48,790	61,656	👁	moses <div>↓ 🔗</div>	txt en <div>↓ 🔗</div>
ELRC-3086-wikipedia_health v1	4,274	91,679	126,414	👁	moses <div>↓ 🔗</div>	txt en <div>↓ 🔗</div>
ELRC_2922 v1	4,273	91,693	126,417	👁	moses <div>↓ 🔗</div>	txt en <div>↓ 🔗</div>
GNOME v1	4,140	44,646	700,000	👁	moses <div>↓ 🔗</div>	txt en <div>↓ 🔗</div>

Hundreds of dataset for text translation task

Thanks for watching