Java AI - 机器翻译

网址：www.aias.top

介绍

笔记简介：
- 面向对象：深度学习初学者
- 依赖课程：**线性代数，统计概率**，优化理论，图论，离散数学，微积分，信息论

知乎专栏：
https://zhuanlan.zhihu.com/p/693738275

Github & Gitee 地址：
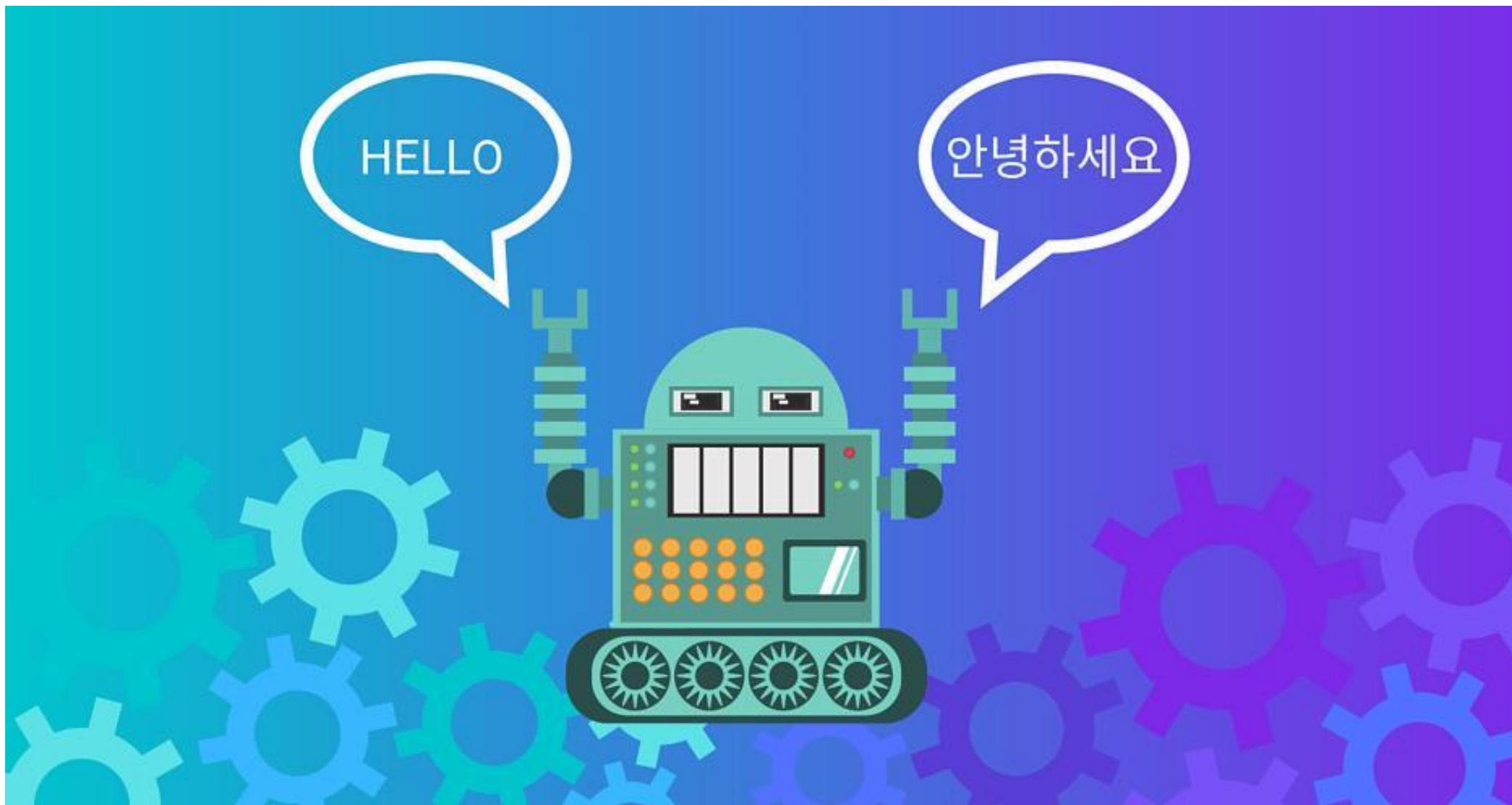https://github.com/mymagicpower/AIAS/tree/main/deep_learning

https://gitee.com/mymagicpower/AIAS/tree/main/deep_learning

* 版权声明：
- 仅限用于个人学习
- 禁止用于任何商业用途

机器翻译

202种语言互相翻译 - SDK

主要特性
- 支持202种语言互相翻译。
- 支持 CPU / GPU

```java
SearchConfig config = new SearchConfig();
// 设置输出文字的最大长度
config.setMaxSeqLength(128);
// 设置源语言: 中文 "zho_Hans": 256200
config.setSrcLangId(256200);
// 设置目标语言: 英文 "eng_Latn": 256047
config.setForcedBosTokenId(256047);

// 输入文字
String input = "智利北部的丘基卡马塔矿是世界上最大的露天矿之一，长约4公里，宽3公里，深1公里。";

String modelPath = "models/";
String cpuModelName = "traced_translation_cpu.pt";
String gpuModelName = "traced_translation_gpu.pt";
try (NllbModel nllbModel = new NllbModel(config, modelPath, cpuModelName, Device.cp

    System.setProperty("ai.djl.pytorch.graph_optimizer", "false");

    // 运行模型，获取翻译结果
    String result = nllbModel.translate(input);

    logger.info("{}", result);
```


Language	FLORES-200 code	Language Id
Acehnese (Arabic script)	ace_Arab	256001
Acehnese (Latin script)	ace_Latn	256002
Mesopotamian Arabic	acm_Arab	256003
Ta'izzi-Adeni Arabic	acq_Arab	256004
Tunisian Arabic	aeb_Arab	256005
Afrikaans	afr_Latn	256006
South Levantine Arabic	ajp_Arab	256007
Akan	aka_Latn	256008
Amharic	amh_Ethi	256009
North Levantine Arabic	apc_Arab	256010
Modern Standard Arabic	arb_Arab	256011
Najdi Arabic	ars_Arab	256012
Moroccan Arabic	ary_Arab	256013
Egyptian Arabic	arz_Arab	256014
Assamese	asm_Beng	256015
Asturian	ast_Latn	256016
Awadhi	awa_Deva	256017
Central Aymara	ayr_Latn	256018
South Azerbaijani	azb_Arab	256019
North Azerbaijani	azj_Latn	256020
Bashkir	bak_Cyrl	256021
Bambara	bam_Latn	256022
Balinese	ban_Latn	256023
Belarusian	bel_Cyrl	256024
Bemba	bem_Latn	256025

中英互相翻译 - SDK

主要特性
- 可以进行英语和中文之间的翻译
- 支持 CPU / GPU

```java
SearchConfig config = new SearchConfig();
config.setMaxSeqLength(128);

String modelPath = "models/opus-mt-zh-en/";

String input = "智利北部的丘基卡马塔矿是世界上最大的露天矿之一，长约4公里，宽3公里，深1公里。";

try (TranslationModel translationModel = new TranslationModel(config, modelPath, poolSize: 4, Device.cpu());
) {
    System.setProperty("ai.djl.pytorch.graph_optimizer", "false");

    long start = System.currentTimeMillis();
    String result = translationModel.translate(input);
    long end = System.currentTimeMillis();
    logger.info("Time: {}", (end - start));
    logger.info("{}", result);
```

输出翻译结果：

```
[INFO ] -  The Chuquicamata mine in northern Chile is one of the world's largest, about 4 km long, 3 km wide and 1 km deep.
```

网址

www.aias.top