

## گزارش پروژه‌ی ترجمه‌ی ماشینی

تصحیح انواع خطا (گرامری و غیره) توسط دیتاست موازی

دیتاست از زبان‌آموزان کمبریج جمع‌آوری شده است که خطای آن توسط متخصص تصحیح شده است. این دیتاست شامل متن‌های نوشته شده توسط زبان‌آموزانی است که زبان انگلیسی به عنوان زبان دوم آن‌ها محسوب می‌شود.

این دیتاست شامل 1244 عدد متن می‌باشد که پس از پردازش به ۱۶۰۶۸ پاراگراف تبدیل می‌شود.  
نمونه:

```
<coded_answer>
<p>Dear Mr Ryan<NS type="RP"><i>.</i></c></NS></p>
<p>Thanks for <NS type="DD"><i>you</i></c></NS> your<NS type="MD"><i>the</i></c></NS> information you need and ask some questions.</p>
<p>I <NS type="TV"><i>could</i></c></NS> only travel <NS type="RT"><i>on</i></c></NS> July. As you know, I am a student and the nearest holiday is <NS type="MD"><i>the</i></c></NS> summer holiday. But I have booked a flight <NS type="UT"><i>to</i></c></NS> home at the beginning of <NS type="S"><i>August</i></c></NS>. And also I would like to go <NS type="RT"><i>on</i></c></NS> summer.</p>
<p><NS type="MT"><i>The</i></c></NS> accommodation<NS type="MP"><i>,</i></c></NS> I would like to <NS type="RV"><i>live</i></c></NS> in tents<NS type="RP"><i>.</i></c></NS> Because<NS type="TV"><i>never</i></c></NS> I <NS type="RV"><i>live</i></c></NS> in tents<NS type="MD"><i>a</i></c></NS> <NS type="AGN"><i>tents</i></c></NS> before. I think it <NS type="TV"><i>is</i></c></NS> great and I want to try it.</p>
<p>I like doing sports. I would like to play basketball and golf when I am at the Camp. I play basketball a lot and I am a member of our college <NS type="RN"><i>term</i></c></NS>. But I am not very good at golf.</p>
<p>And also I want to ask some questions. What clothes should I <NS type="TV"><i>take</i></c></NS>? How much money should I <NS type="TV"><i>take</i></c></NS>? And how <NS type="TV"><i>could</i></c></NS> we meet at the airport? I am looking forward <NS type="MT"><i>to</i></c></NS> your reply.</p>
<p>Yours sincerely.</p>
</coded_answer>
```

برای جنریت دیتاست موازی ابتدا باید فرمت xml دیتاست را خوانده و از روی آن یک دیتاست برای موارد خطا (مبدا) و یک دیتاست برای متن درست شده (مقصد) تولید شود.

برای بررسی عملکرد از امتیاز bleu استفاده کردیم ولی به دلیل اینکه خطاهای تصحیح شده عموماً در حد یکی دو کلمه هستند، این امتیاز به صورت مصنوعی خیلی خوب به دست می‌آید، که در آزمایش ما برابر با 0.80 شد.

برای ساخت مدل از کیت OpenNMT استفاده می‌کنیم. ساده ترین مدل این کیت یک LSTM encoder decoder است. برای بهبود نتایج نیز از word-embedding های glove نسخه‌ی ۱۰۰ بعدی آن استفاده کردیم.

نمونه ترجمه: (جمله مبدا، ترجمه‌ی درست، ترجمه ماشین)

1-

I would be grateful if you could let me know the above information as soon as possible . Thank you once again for giving me good news !  
I would be grateful if you could let me have the above information as soon as possible . Thank you once again for giving me this good news !  
I would be grateful if you could let me know the above information as soon as possible . Thank you once again for giving me a good news !

2-

Dear Helen Ryan .  
Dear Helen Ryan ,  
Dear Helen Ryan ,

3-

With the help of some modern technology our daily life would not so difficult .  
With the help of some modern technology our daily lives would not be so difficult .  
With the help of modern technology our daily life would not be so difficult .

4-

Dear Sir or Mam ,  
Dear Sir or Madam ,  
Dear Sir or Madam ,

5-

Modern technology have change my daily life in many aspects .  
Modern technology has changed my daily life in many ways .  
Modern technology has changed my daily life in many aspects .

6-

In the opposite , shopping can be disenjoyable .  
On the other hand , shopping can be unenjoyable .  
In the opposite , shopping can be scary .