



## به نام خدا تمرین کامپیوتری امتیازی

پرواز زبان‌های طبیعی - بهار ۱۴۰۳

دانشکده مهندسی برق و کامپیوتر

طراح تمرین: **علی افتخاری**

مهلت تحویل: ۱۳ تیر ۱۴۰۳

مدرس: دکتر هشام فیلی

### مقدمه

امروزه با توجه به پیشرفت‌های روز افزون هوش مصنوعی مدل‌های متفاوت با کارکردهای فراوان ساخته شده‌اند که یکی از این مدل‌ها، مدل [TTS](#)<sup>1</sup> می‌باشد که می‌تواند داده‌های متنی را به صوت تبدیل کند، این مدل امروزه برای ایجاد دستیار صوتی در دستگاه‌های هوشمند استفاده می‌شود.

برای ساخت این مدل باید دیتای مناسب آن وجود داشته باشد که بخشی از آن به صورت صوت و بخش دیگر متن می‌باشد.

در بخش دیتای متنی برای ساخت این مدل باید به گونه‌ای باشد که آواها در آن مشخص شده باشند یعنی به صورت فونیم<sup>2</sup> باشند. حال به دلیل آنکه به حجم بالایی از این دیتا نیاز داریم، باید مدل [G2P](#)<sup>3</sup> طراحی کنیم که این نوع دیتا را تولید کند. برای ساخت همچنین مدلی باز به دیتای متن فونیم و جمله خام آن نیاز خواهیم داشت.

<sup>1</sup> Text-To-Speech

<sup>2</sup> Phoneme

<sup>3</sup> Grapheme-to-Phoneme

## توضیحات

در این نوع دیتا باید جملات فارسی که به شما داده می شود را به صورتی که در مثال آورده شده بنویسید:

- طرح افتتاح حساب پس انداز زنان کشور توسط بانک کشاورزی اجرا میشود.

**tarhe eftetAhe hesAbe pas'andAze zanAne keSvar tavassote bAnke keSAvarzi ejra miSavad.**

- کلیات طرح بازنشتگی زود هنگام زنان تصویب شد.

**kolliyAte tarhe bAzneSastegiye zudhangAme zanAn tasvib Sod.**

- یک دختر ایرانی، برنده جایزه ویژه آکادمی های پزشکی روسیه شد.

**yek doxtare irAni , barandeye jAyezeye viZeye AkAdemihAye pezeSkiye rusiye Sod.**

اگر کمی دقت کنید متوجه خواهید شد که جملات فونیم خیلی مانند تایپ "فینگلیش" می باشد که شاید روزمره از آن استفاده می کنید!

همانطور که مشاهده می کنید کمی تفاوت را در نوع فونیم این جملات با تایپ "فینگلیش" وجود دارد که از شما انتظار می رود تا جملات فارسی که به شما داده می شود را مانند نمونه به متن متناظر فونیم آن تبدیل کنید.

برای راحتی کار شما حروف متناظر هر حرف فارسی قرار داده شده:

A	آ
a	اَ
S	ش
s	س
Z	ژ
z	ز، ذ، ض، ظ
u	او
o	اُ
x	خ
باید دوبار حرف متناظر آن تکرار شود.	ّ (تشدید)

h	ح
C	چ
p	پ
t	ت
k	ک
f	ف
b	ب
d	د
g	گ
q	ق، غ
v	و
j	ج
r	ر
m	م
n	ن
l	ل
y	ی
i	ای
e	اِ
a یا A یا e (بستگی به حرف صدا دار چسبیده به آن را دارد)	ع

## نکات تکمیلی

- استفاده از کدهای آماده برای پیاده‌سازی این مبحث (جستجو شده در اینترنت و ...)، در صورتی که منبع ذکر شود مجاز می‌باشد.
- از شما انتظار می‌رود که حداقل ۲۵۰ جمله فارسی را به صورت فونیم آن بنویسید. در صورتی که کمتر از این مقدار را انجام دهید نمره امتیازی به شما کمتر تعلق می‌گیرد.
- تمامی جملات شما بررسی شده و اگر مقدار زیادی از آن به درستی نوشته نشده باشد نمره‌ای به شما تعلق نخواهد گرفت.
- خروجی شما باید یک فایل CSV از جملات فارسی و جملات فونیم متناظر آن مانند نمونه باشد.
- در صورتی که مایل به انجام این تمرین دارید در گروه تلگرام زیر عضو شوید تا جملات فارسی در اختیار شما گذاشته شود:

[https://t.me/+em0hMH0\\_rKM1MjE0](https://t.me/+em0hMH0_rKM1MjE0)

[ali.eftekhari79uni@gmail.com](mailto:ali.eftekhari79uni@gmail.com)

در صورت نیاز با بنده در ارتباط باشید: