



درس پردازش زبان های طبیعی
تمرین رایانه ای ششم

ساخت ربات گفت و گو با ابزار Rasa

تدریس یاران	پرسش یک	نام و نام خانوادگی	میلاد رئیسی
	پرسش دو	نام و نام خانوادگی	زهرا پیروی
		رایانامه	MiladReisi97@gmail.com
		رایانامه	ZahraPeiravi20@gmail.com
مهلت ارسال تمرین		۱۴۰۲.۰۳.۳۱	

فهرست

قوانین	ب
مقدمه	۱
ربات پاسخ‌گو به پرسش‌های پرتکرار	۱
پرسش ۱. ربات پاسخ‌گو به پرسش‌های پرتکرار	۳
۱-۱. آماده‌سازی داده‌ها	۴
۱-۱-۱. آماده‌سازی داده‌ها به صورت دستی	۴
۱-۱-۲. آماده‌سازی داده‌ها به صورت خودکار	۵
۲-۱. انجام تنظیمات و آموزش ربات	۵
۳-۱. ارزیابی ربات	۶
۴-۱. ابزار گفت‌وگوی تحت وب	۶
پرسش ۲. استخراج مقادیر ارزش‌ها	۸
سناریو اول: استخراج مقادیر ارزش‌های نام دانشجو و شماره دانشجویی	۸
سناریو دوم: استخراج مقادیر ارزش‌های نام فیلم و سال ساخت فیلم	۹
سناریو سوم: استخراج مقادیر ارزش‌های نام میوه و وزن (کیلوگرم/گرم) آن	۱۰
سناریو چهارم: استخراج مقادیر ارزش‌های برند تلویزیون و ابعاد آن	۱۰
سناریو پنجم: استخراج مقادیر ارزش‌های نام شهری محل زندگی و شماره‌ی همراه کاربر	۱۱
سناریو ششم: استخراج مقادیر ارزش‌های نام شهر مقصد و تاریخ پرواز	۱۲
۱-۲. آماده‌سازی دادگان و آموزش مدل	۱۳
۲-۲. پیاده‌سازی و تحلیل نتایج	۱۴

قبل از پاسخ دادن به پرسش‌ها، موارد زیر را با دقت مطالعه نمایید:

- پرسش دو امتیازی بوده و مقدار تاثیر نمره‌ی آن بر روی نمرات شما بعدا تعیین و اعلام خواهد شد. بنابراین با انجام پرسش یک نمره‌ی کامل تمرین را دریافت خواهید کرد.
- کیفیت گزارش شما در فرآیند تصحیح از اهمیت ویژه‌ای برخوردار است؛ بنابراین، لطفا تمامی نکات و فرض‌هایی را که در پیاده‌سازی‌ها و محاسبات خود در نظر می‌گیرید در گزارش ذکر کنید.
- به تمرین‌هایی که به صورت کاغذی تحویل داده شوند یا به صورت عکس در سایت بارگذاری شوند، ترتیب اثری داده نخواهد شد.
- تحلیل نتایج الزامی می‌باشد، حتی اگر در صورت پرسش اشاره‌ای به آن نشده باشد.
- دستیاران آموزشی ملزم به اجرا کردن کدهای شما نیستند؛ بنابراین، هرگونه نتیجه و یا تحلیلی که در صورت پرسش از شما خواسته شده را به طور واضح و کامل در گزارش بیاورید. در صورت عدم رعایت این مورد، بدیهی است که از نمره تمرین کسر می‌شود.
- کدهای نوشته شده برای هر بخش را با نام مناسب مشخص کرده و به همراه گزارش تکلیف ارسال کنید. همه‌ی کدهای پیوست گزارش بایستی قابلیت اجرای دوباره داشته باشند. در صورتی که برای اجرا مجدد آنها نیاز به تنظیمات خاصی می‌باشد بایستی تنظیمات مورد نیاز را نیز در گزارش خود ذکر کنید.
- توجه کنید این تمرین باید به صورت تک نفره انجام شود و پاسخ‌های ارائه شده باید نتیجه فعالیت فرد نویسنده باشد (هم‌فکری و به اتفاق هم نوشتن تمرین نیز ممنوع است). در صورت مشاهده تقلب به همه افراد مشارکت کننده، نمره تمرین صفر می‌گردد.
- تنها زبان برنامه نویسی مجاز Python است.
- لطفا گزارش، کدها و سایر ضمایم را به در یک پوشه با نام زیر قرار داده و آن را فشرده سازید، سپس در سامانه‌ی Elearn بارگذاری نمایید:

NLP_CA6_[StudentID].zip

(مثال: NLP_CA6_810101101.zip)

به طور سنتی بسیاری از شرکت‌ها برای فراهم آوردن خدمات پشتیبانی از طریق گفت‌وگوی زنده، کارشناسانی تربیت می‌کنند تا بتوانند هر چه سریع‌تر و باکیفیت‌تر پاسخ‌گوی مشتریان باشند؛ با این حال، با توجه به این که تعداد مشتریان زیاد و همواره نیز رو به افزایش است، در عمل معطلی‌های طولانی تا لحظه‌ای اتصال به کارشناس مشاهده و حتی پس از آن نیز، هنگام گفت‌وگو، زمان زیادی در انتظار نوشته شدن پاسخ توسط کارشناس تلف می‌شود که این موارد سبب نارضایتی مشتریان می‌گردد. علاوه بر این‌ها و صرف نظر از هزینه‌های زیادی که تربیت و حفظ کارشناسان دارد، یک کارشناس قادر به پاسخ‌گویی شبانه‌روزی نیز نیست. در نتیجه به عنوان راه حل، استفاده از یک سامانه‌ی نرم افزاری هوشمند^۱ پیشنهاد می‌شود که بتواند به زبان طبیعی^۲ با مشتریان گفت‌وگو کند و با استفاده از دانش درونی یا بیرونی (مانند پایگاه دانش و ...) سریع‌تر، بدون معطلی و شبانه‌روزی، پاسخ‌گوی درخواست‌ها و پرسش‌های مشتریان باشد. به این سامانه‌ی نرم افزاری هوشمند که از هوش مصنوعی بهره می‌گیرد تا بتواند به زبان طبیعی با انسان گفت‌وگو کند، **ربات گفت‌وگو**^۳ می‌گویند.

ربات پاسخ‌گو به پرسش‌های پرتکرار^۴

امروزه استفاده از ربات‌های گفت‌وگو در زمینه‌ی خدمات مشتریان^۵ در شرکت‌های بزرگ خارجی (مانند آمازون^۶) و داخلی (مانند همراه اول) به سرعت رو به افزایش است چرا که علاوه بر مزایایی که پیش‌تر گفته شد، هزینه‌های پشتیبانی و خدمات شرکت‌ها را نیز بسیار کاهش می‌دهند. معمولاً شرکت‌ها بخشی به نام مرکز تماس^۷ دارند که در آن به درخواست‌های گوناگون مشتریان پاسخ می‌دهند. بخش قابل توجهی از درخواست‌های مشتریان را پرسش‌های خاص و مشخصی تشکیل می‌دهد که از آن با عنوان پرسش‌های پرتکرار^۸ یاد می‌شود. به طور سنتی مدیران مراکز تماس و پشتیبانی شرکت‌ها، این پرسش‌های پرتکرار را جمع‌آوری و دسته‌بندی نموده و همراه با پاسخ آن‌ها در اختیار کارشناسان پاسخ‌گو و یا بر روی پایگاه اینترنتی^۹ خود قرار می‌دهند. اما امروزه بسیاری از شرکت‌ها علاوه بر این‌ها، در تلاش‌اند تا از ربات‌های

Intelligent software system^۱

Natural Language^۲

Chatbot^۳

FAQ Chatbot^۴

Customer service^۵

Amazon^۶

Call Center^۷

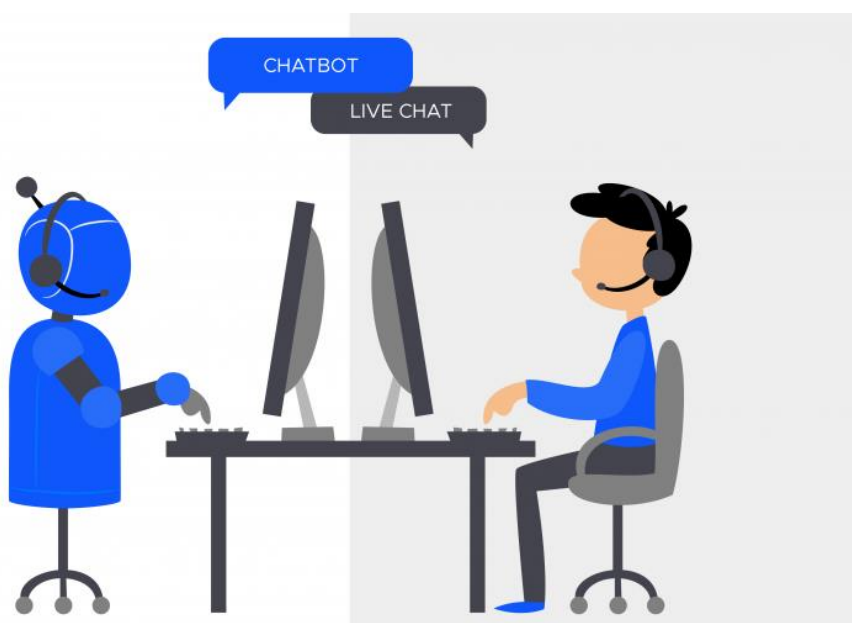
Frequently Asked Questions (FAQ)^۸

Website^۹

گفت‌وگو نیز در این زمینه استفاده کنند. ربات گفت‌وگویی که به طور خاص، تنها پاسخ‌گوی این بخش از درخواست‌های کاربر باشد، ربات پاسخ‌گو به پرسش‌های پرتکرار نامیده می‌شود.

پرسش ۱. ربات پاسخ‌گو به پرسش‌های پرتکرار^۱

در این پرسش با ربات پاسخ‌گو به پرسش‌های پرتکرار آشنا خواهید شد و یک نمونه از آن را با ابزار **Rasa** پیاده‌سازی خواهید کرد. پرسش و پاسخ‌های پرتکراری که در نظر گرفته شده، پرسش و پاسخ‌های پرتکرار **شرکت ارتباطات سیار ایران، همراه اول** در زمینه‌ی **اینترنت** است. شکل ۱ تصویری نمادین از گفت‌وگوی یک انسان با یک ربات گفت‌وگو را نشان می‌دهد.



شکل ۱. تصویری نمادین از گفت‌وگوی انسان با ربات گفت‌وگو

با توجه به آموزشی که در قالب فیلم به پیوست فرستاده شده، پس از نصب **Rasa**، یک پروژه‌ی ابتدایی از آن ایجاد خواهید کرد. سپس اقدام به شخصی‌سازی بخش‌های مختلف پروژه و تغییر آن به شکلی که تبدیل به ربات پاسخ‌گو به پرسش‌های پرتکرار شود خواهید نمود و در نهایت آن را آموزش خواهید داد تا ربات ساخته شود. برای ساخت ربات پاسخ‌گو به پرسش‌های پرتکرار با ابزار **Rasa** می‌توان دو رویکرد کلی داشت:

^۱ FAQ Bot

۱. هر جفت پرسش و پاسخ، یک کلاس^۱ یا اندیشه^۲ی جداگانه است.

۲. یک کلاس یا اندیشه‌ی واحد به نام پرسش و پاسخ‌های پرتکرار داریم و هر جفت پرسش و پاسخ، یک زیرکلاس^۳ یا زیراندیشه^۴ از کلاس یا اندیشه‌ی مادر، یعنی اندیشه‌ی پرسش و پاسخ‌های پرتکرار هستند.

در این تمرین هر دو نوع رویکرد را پیاده‌سازی خواهید کرد. تفاوت دو رویکرد در نحوه آماده‌سازی داده‌ها و انجام تنظیمات **Rasa** است که این تفاوت‌ها را باید خودتان از **پایگاه اینترنتی مستندات Rasa** مطالعه نمایید. پس دقت کنید که در مجموع، بخش‌های ۱-۱ تا ۳-۱ را باید یک بار برای رویکرد اول و یک بار برای رویکرد دوم انجام دهید. (پیاده‌سازی این سه بخش هر رویکرد ۵۰ نمره خواهد داشت)

۱-۱. آماده‌سازی داده‌ها

داده‌ها در قالب جدولی با نام **MCI_Internet_TrainData** و با پسوند **.xlsx** به پیوست فرستاده شده‌اند. هر سطر در این جدول مربوط به یک اندیشه است که در آن هر اندیشه یک پاسخ و تعدادی نمونه پرسش دارد. مشابه با آنچه در یادگیری ماشین به آن عمل طبقه‌بندی^۵ می‌گویند، در این جا نیز ابتدا ربات با نمونه پرسش‌های هر اندیشه آموزش می‌بیند و یک مدل شبکه‌ی عصبی تولید می‌کند. پس از آن، در عمل و هنگام کار و اجرا وقتی کاربر چیزی از ربات می‌پرسد، ربات اندیشه‌ی او را بر اساس مدل آموزش دیده مشخص می‌کند، سپس پاسخ مربوط به همان اندیشه را به کاربر برمی‌گرداند.

در این بخش باید داده‌ها را در پرونده‌های مخصوص به خود در پروژه‌ی **Rasa** قرار دهید. پاسخ‌ها در پرونده‌ی **domain.yml** و پرسش‌های نمونه در پرونده‌ی **nlu.yml** (که در پوشه‌ی **data** قرار دارد) قرار می‌گیرند. برای این کار دو راه پیش رو دارید:

۱-۱-۱. آماده‌سازی داده‌ها به صورت دستی

(۱۰ نمره)

به صورت دستی تک تک داده‌ها را در قالبی مشابه با آنچه در پرونده‌های **nlu.yml** و **domain.yml** مشاهده می‌کنید قرار دهید.

¹ Class

² Intent

³ Sub-Class

⁴ Sub-Intent

⁵ Classification

۱-۲. آماده سازی داده‌ها به صورت خودکار

(۱۲.۵ نمره) [۲.۵ نمره امتیازی و اضافه برای نوشتن کد هر رویکرد در نظر گرفته شده است. پس در مجموع ۵ نمره اضافه به ۱۰۰

نمره‌ی تمرین اضافه خواهد شد.]

کدی به زبان **Python** بنویسید که ابتدا داده‌ها را از پرونده‌ی **MCI_Internet_TrainData** بخواند، سپس آن‌ها را با ساختار صحیح، در پرونده‌های **nlu.yml** و **domain.yml** قرار دهد.

۱-۲. انجام تنظیمات و آموزش^۱ ربات

(۳۰ نمره)

ابزار **Rasa** برای ساخت ربات‌های گوناگونی مورد استفاده قرار می‌گیرد و هر ربات نیز تنظیماتی مخصوص به خود دارد. در این بخش شما باید با تغییر در پرونده‌ی **config.yml** تنظیمات مورد نیاز برای ساخت ربات پاسخ‌گو به پرسش‌های پرتکرار را متناسب با رویکردی که در حال اجرای آن هستید برگزینید. در این تمرین قصد داریم با عملکرد مدل‌های زبانی بر پایه‌ی **Bert** آشنا شوید. بنابراین باید ابتدا مدل‌های **ParsBert** و **LaBSE** را بیازمایید و مواردی که در بخش ۱-۳ گفته شده را برای هر یک گزارش کنید. سپس هر یک از این مدل‌ها که بهتر بود را انتخاب کنید و تاثیر تعداد دورهای آموزش^۲ را روی آن بررسی کنید. به این صورت که یک بار با ۵۰ دور، یک بار با ۱۰۰ دور و یک بار با ۲۰۰ دور آموزش را انجام دهید و در نهایت حالت بهینه را گزارش کنید. هر بار پس از انجام تغییرات، از محتوای پرونده‌ی **config.yml** عکس بگیرید و در گزارش خود بیاورید. سپس با توجه به عکس، تغییرات انجام شده را به طور مختصر توضیح دهید و کاربرد و دلیل هر یک از آن‌ها را بنویسید. (بدیهی است بار اول خط به خط تنظیمات را توضیح می‌دهید و در آموزش‌های بعدی فقط تغییرات انجام شده را نسبت به تنظیمات اولیه توضیح خواهید داد. ۳۰ نمره‌ی در نظر گرفته شده برای این بخش مربوط به انجام درست تنظیمات و توضیحاتی که درباره‌ی آن می‌دهید است)

^۱ Train

^۲ Epochs

۱-۳. ارزیابی^۱ ربات

(۱۰ نمره)

با استفاده از دادگانی (نمونه پرسش‌هایی) که در پرونده‌ی **MCI_Internet_TestData** قرار گرفته، ربات را ارزیابی کنید و سپس **ماتریس آشفتگی**^۲ و مقادیر **دقت**^۳ و **امتیاز F1**^۴ را برای هر اندیشه گزارش کنید. (چگونگی انجام ارزیابی با استفاده از دادگانی غیر از دادگان آموزشی را از پایگاه اینترنتی مستندات Rasa مطالعه نمایید.)

توجه: **ماتریس آشفتگی** و مقادیر **دقت** و **امتیاز F1** را **Rasa** خودش محاسبه می‌کند و در اختیارتان قرار می‌دهد و شما تنها کافی است آن‌ها در گزارش قرار دهید.

۱-۴. ابزار گفت‌وگوی تحت وب

(بدون نمره و دلخواه)

به پیوست تمرین یک پوشه به نام **WebChat** قرار داده شده. این پوشه را به صورت کامل به پوشه‌ی پروژه‌ی ربات خود انتقال دهید. پس از آموزش و تکمیل ربات خود، برای گفت‌وگو با آن دو راه پیش رو دارید:

۱. استفاده از خط فرمان

۲. استفاده از ابزار گفت‌وگوی تحت وب

حالت اول که حالت ساده و ابتدایی است، با اجرای دستور زیر در خط فرمان امکان‌پذیر است:

rasa shell

حالت دوم استفاده از یک رابط تصویری تحت وب است. برای استفاده از آن کافی است پس از ساخت ربات، در خط فرمان دستور زیر را اجرا کنید:

rasa run -m models --enable-api --cors "" --debug*

¹ Test

² Confusion Matrix

³ Accuracy

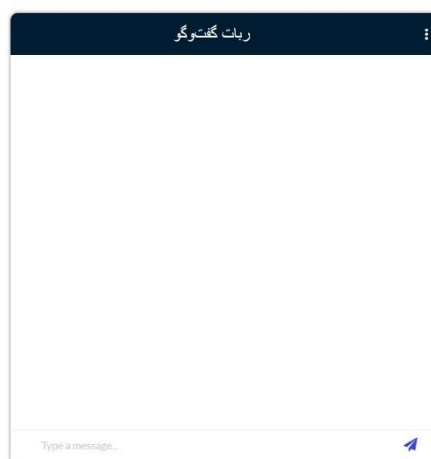
⁴ F1 Score

سپس وارد پوشه‌ی **WebChat** شده و پرونده‌ی **index.html** را باز کنید. سپس همان‌طور که در شکل ۲ نشان داده شده، گوشه‌ی پایین-راست تصویر را لمس کنید تا صفحه‌ی گفت‌وگو با ربات نمایان شود.



شکل ۲. صفحه‌ی ابتدایی پرونده‌ی **index.html**

شکل ۳ نمونه‌ای از صفحه‌ی گفت‌وگو با ربات را نشان می‌دهد. حال می‌توانید به راحتی با ربات خود گفت‌وگو کنید.



شکل ۳. نمونه‌ای از صفحه‌ی گفت‌وگو با ربات

پرسش ۲. استخراج مقادیر ارزش‌ها^۱

در این پرسش با ارزش‌ها و چگونگی استخراج مقادیر آن‌ها از جملات آشنا خواهید شد. ارزش^۲، تعریفی مشخص از یک ویژگی است که می‌تواند مقادیر مختلفی داشته باشد و استخراج مقادیر مختلف آن در متن برای ما اهمیت دارد. برای مثال، در جمله‌ی «من می‌خواهم بلیت هواپیما تهیه کنم»، **نوع وسیله‌ی نقلیه** یک ارزش است که مقدار آن در این جمله **هواپیما** است. در علم پردازش زبان‌های طبیعی، این جمله به صورت «من می‌خواهم بلیت [هواپیما] (وسیله نقلیه) تهیه کنم» نشانه‌گذاری^۳ می‌شود که در آن هم ارزش و هم مقدار آن مشخص شده‌اند.

همانند پرسش قبل در این پرسش نیز از ابزار **Rasa** استفاده خواهید کرد. هدف این پرسش استخراج مقادیر دو نوع گوناگون ارزش است؛ یکی ارزشی که مقادیر آن با عبارت منظم قابل استخراج است و دیگری ارزشی که برای استخراج مقادیر آن باید از ابزارهای پیشرفته‌تری استفاده کرد. برای اولی از ابزار **Regex** **entity extractor** و برای دومی از ابزار **CRF entity extractor** استفاده خواهید نمود. ۶ سناریو در نظر گرفته شده که به هر یک از دانشجویان با توجه به **باقیمانده ۳ رقم آخر شماره دانشجویی خود بر ۶ به علاوه‌ی ۱** یکی از سناریوهای زیر اختصاص می‌یابد.

مثال:

شماره دانشجویی: ۸۱۰۱۰۱۰۶۲

سناریوی اختصاص داده شده: سناریوی سوم

$$062 \% 6 = 2$$

$$2 + 1 = 3$$

سناریو اول: استخراج مقادیر ارزش‌های نام دانشجو و شماره دانشجویی

کاربر درخواست انتخاب واحد دارد. ربات از او نام و شماره دانشجویی‌اش را درخواست می‌کند. پس از این که کاربر نام و شماره دانشجویی‌اش را وارد کرد، ربات باید مقادیر ارزش‌های نام و شماره دانشجویی را از گفته‌ی

¹ Slot values

² Slot

³ Annotate

کاربر استخراج کرده و در یک پیام به کاربر نشان دهد و تایید کاربر را از درست بودن اطلاعات استخراج شده دریافت کند.

مثال:

- کاربر: می‌خواهم انتخاب واحد کنم
- ربات: لطفا نام و شماره دانشجویی خود را بفرستید
- کاربر: اسم من علی حسینی است و شماره دانشجویی‌ام ۸۱۰۱۰۰۲۰۰ است
- ربات: اطلاعات شما (نام "علی حسینی" با شماره دانشجویی "۸۱۰۱۰۰۲۰۰") دریافت شد. آیا از صحت اطلاعات وارد شده اطمینان دارید؟
- در صورت درست بودن اطلاعات:
 - کاربر: بله
 - ربات: در حال ورود به صفحه انتخاب واحد...
- در صورت نادرست بودن اطلاعات:
 - کاربر: خیر
 - ربات: لطفا یک بار دیگر اطلاعات خود را وارد کنید.

سناریو دوم: استخراج مقادیر ارزش‌های نام فیلم و سال ساخت فیلم

کاربر از ربات درخواست تماشای فیلم دارد. ربات از کاربر می‌خواهد نام فیلم و سال ساخت فیلم را وارد کند. کاربر نام فیلم مورد نظر خود و سال ساخت فیلم را می‌گوید. ربات باید مقادیر ارزش‌های نام فیلم و سال ساخت فیلم را از گفته‌ی کاربر استخراج کرده و در یک پیام به او نشان دهد و تایید کاربر را از درست بودن اطلاعات استخراج شده دریافت کند.

مثال:

- کاربر: می‌خواهم یک فیلم تماشا کنم
- ربات: لطفا نام فیلم و سال ساخت فیلم مورد نظر خود را بگویید.
- کاربر: فیلم شبی که ماه کامل شد که در سال ۱۳۹۷ ساخته شده
- ربات: آیا شما می‌خواهید فیلم "شبی که ماه کامل شد" ساخته شده در سال "۱۳۹۷" را تماشا کنید؟
- در صورت درست بودن اطلاعات:

- کاربر: بله
- ربات: در حال جستجو فیلم مورد نظر...
- در صورت نادرست بودن اطلاعات:
- کاربر: خیر
- ربات: لطفا یک بار دیگر اطلاعات فیلم مورد نظر خود را وارد کنید.

سناریو سوم: استخراج مقادیر ارزش‌های نام میوه و وزن (کیلوگرم/گرم) آن

کاربر از ربات درخواست ثبت خرید میوه را دارد. ربات از کاربر می‌خواهد نام میوه دلخواه و وزن آن را وارد کند. کاربر نام و وزن را می‌گوید. ربات باید مقادیر ارزش‌های نام میوه و وزن آن را از گفته‌ی کاربر استخراج کرده و در یک پیام به او نشان دهد و تایید کاربر را از درست بودن اطلاعات استخراج شده دریافت کند.

مثال:

- کاربر: میخوام میوه بخرم
- ربات: برای تکمیل خرید خود لطفا نام میوه مدنظر و مقداری که از آن می‌خواهید خریداری کنید را وارد کنید.
- کاربر: میخوام ۱ کیلو سیب بخرم
- ربات: آیا می‌خواهید "۱ کیلو" "سیب" را به سبد خرید شما اضافه کنم؟
- در صورت درست بودن اطلاعات:
- کاربر: بله
- ربات: به سبد خرید شما اضافه شد
- در صورت نادرست بودن اطلاعات:
- کاربر: خیر
- ربات: لطفا یک بار دیگر اطلاعات مورد نیاز را بگویید

سناریو چهارم: استخراج مقادیر ارزش‌های برند تلویزیون و ابعاد آن

کاربر از ربات درخواست ثبت خرید تلویزیون را دارد. ربات از کاربر می‌خواهد برند تلویزیون دلخواه و ابعاد آن را وارد کند. کاربر برند و ابعاد را می‌گوید. ربات باید مقادیر ارزش‌های نام برند و ابعاد تلویزیون را از

گفته‌ی کاربر استخراج کرده و در یک پیام به او نشان دهد و تایید کاربر را از درست بودن اطلاعات استخراج شده دریافت کند.

مثال:

- کاربر: میخوام تلویزیون بخرم
- ربات: برای تکمیل خرید خود لطفا نام برند تلویزیون مدنظر و ابعاد آن را وارد کنید.
- کاربر: میخوام تلویزیون سامسونگ ۵۵ اینچی بخرم
- ربات: آیا میخواهید تلویزیون "سامسونگ" "۵۵" اینچی را به سبد خرید شما اضافه کنم؟
- در صورت درست بودن اطلاعات:
 - کاربر: بله
 - ربات: به سبد خرید شما اضافه شد
- در صورت نادرست بودن اطلاعات:
 - کاربر: خیر
 - ربات: لطفا یک بار دیگر اطلاعات مورد نیاز را بگویید

سناریو پنجم: استخراج مقادیر ارزش‌های نام شهری محل زندگی و شماره‌ی همراه کاربر

کاربر از ربات درخواست ثبت عضویت در یک سازمان را دارد. ربات از کاربر می‌خواهد نام شهری که فرد در آنجا زندگی می‌کند و شماره همراه خود را وارد کند. کاربر نام شهر محل سکونت و شماره‌ی تلفن همراه خود را می‌گوید. ربات باید مقادیر ارزش‌های نام شهر و شماره همراه فرد را از گفته‌ی کاربر استخراج کرده و در یک پیام به او نشان دهد و تایید کاربر را از درست بودن اطلاعات استخراج شده دریافت کند.

مثال:

- کاربر: میخوام عضو سازمان بشم
- ربات: برای تکمیل عضویت لطفا نام و شماره همراه خود را وارد کنید.
- کاربر: علی حسینی هستم و شماره همراه من ۰۹۱۲۱۲۳۴۵۶۷ است
- ربات: آیا اطلاعات وارد شده شما با نام "علی حسینی" با شماره همراه "۰۹۱۲۱۲۳۴۵۶۷" صحیح است؟
- در صورت درست بودن اطلاعات:

- کاربر: بله
- ربات: شما به عضویت سازمان در آمدید
- در صورت نادرست بودن اطلاعات:
- کاربر: خیر
- ربات: لطفا یک بار دیگر اطلاعات مورد نیاز را بگویید

سناریو ششم: استخراج مقادیر ارزش‌های نام شهر مقصد و تاریخ پرواز

کاربر از ربات درخواست جستجوی بلیت هواپیما برای سفر به شهری خاص در تاریخی معین را دارد. ربات از کاربر می‌خواهد نام شهری که می‌خواهد به آن جا پرواز کند و تاریخ پرواز مد نظر خود را وارد کند. کاربر نام شهر مقصد و تاریخ را می‌گوید. ربات باید مقادیر ارزش‌های نام شهر مقصد و تاریخ پرواز مورد نظر را از گفته‌ی کاربر استخراج کرده و در یک پیام به او نشان دهد و تایید کاربر را از درست بودن اطلاعات استخراج شده دریافت کند.

مثال:

- کاربر: می‌خواهم به بلیت پرواز بگیرم
- ربات: لطفا نام شهر مقصد و ساعت پرواز خود را بگویید
- کاربر: می‌خواهم به مشهد سفر کنم و دوست دارم تاریخ پروازم ۱۴۰۲/۰۳/۱۹ باشد
- ربات: آیا به دنبال بلیت هواپیما به مقصد "مشهد" در تاریخ "۱۴۰۲/۰۳/۱۹" هستید؟
- در صورت درست بودن اطلاعات:
- کاربر: بله
- ربات: در حال جستجو بلیت برای مقصد و تاریخ مورد نظر شما
- در صورت نادرست بودن اطلاعات:
- کاربر: خیر
- ربات: لطفا یک بار دیگر اطلاعات مورد نیاز را بگویید

۲-۱. آماده سازی دادگان و آموزش مدل

(۲۰ نمره)

همان‌طور که گفته شد برای پیاده سازی سناریوی مربوط به خود باید از ابزار **RASA** استفاده کنید. دقت کنید که شما باید ۲ ارزش را استخراج کنید به طوری که حتما یکی از آن‌ها با استفاده از **Regex entity extractor** و دیگری با استفاده از **CRF entity extractor** استخراج شود. مشابه آن‌چه در پرسش یک انجام شد، باید تعدادی جمله‌ی نمونه به عنوان دادگان آموزشی در پرونده‌ی **nlu.yml** قرار گیرند. بر خلاف پرسش یک که دادگان آموزشی به پیوست فرستاده شده بود، در این پرسش، تهیه‌ی دادگان مناسب به عهده شما قرار دارد. بنابراین باید با توجه به سناریوی اختصاص داده شده به شما، دادگانی را تولید کنید. دقت کنید هر چه نمونه‌های تولیدی بیشتر و متنوع‌تر باشند، عملکرد ربات بهتر خواهد بود.

توجه کنید، همان‌طور که از مثال مربوط به هر سناریو پیداست، ربات باید بتواند چندین اندیشه‌ی متفاوت را به درستی تشخیص دهد. بنابراین شما باید برای هر اندیشه، تعدادی نمونه تهیه کنید و در پرونده‌ی **nlu.yml** قرار دهید.

برای مثال، برای پیاده‌سازی سناریوی ششم، ربات باید بتواند ۴ اندیشه‌ی متفاوت زیر را تشخیص دهد:

۱. مطرح کردن درخواست (میخوام بلیت هواپیما بگیرم)

۲. ارائه‌ی اطلاعات (میخوام به مشهد سفر کنم و دوست دارم تاریخ پروازم ۱۴۰۲/۰۳/۱۹ باشد)

۳. تایید (بله)

۴. عدم تایید (خیر)

شما باید برای هر یک از این اندیشه‌ها مثال‌های مختلفی طرح کنید و به پرونده‌ی **nlu.yml** اضافه کنید. برای مثال، برای اندیشه‌ی **تایید** می‌توان نمونه‌هایی مانند: بله، آره، عاره، تایید می‌کنم، درست است و ... را در نظر بگیرید.

۲-۲. پیاده سازی و تحلیل نتایج

(۸۰ نمره)

موارد زیر را در نظر بگیرید:

- برای پیاده سازی هر تغییری که در فایل **config** خود ایجاد می‌کنید در گزارش خود ذکر کرده و دلیل خود را بیان کنید. (۴۰ نمره)
- بیان کنید کدام یک از ارزش‌های خود را با استفاده از **CRF entity extractor** و کدام یک را با استفاده از **Regex entity extractor** استخراج می‌کنید. (۱۰ نمره)
- شما می‌توانید پس از آموزش مدل، با ربات خود از طریق دستور **rasa shell** گفت‌وگو کنید. از این طریق سناریوی مربوط به خود را اجرا کرده، از گفت‌وگوی انجام شده عکس بگیرید و آن را در گزارش خود بیاورید. (۵ نمره)
- در نهایت، با استفاده از روش **k-fold** مدل را ارزیابی^۱ کرده و نتایج مربوط به استخراج ارزش‌ها و همچنین نتایج مربوط به طبقه‌بندی اندیشه‌ها را گزارش کنید. این نتایج باید شامل **دقت^۲، امتیاز^۳ F1 و ماتریس آشفتگی^۴** باشد. (۱۰ نمره)
- برای استخراج مقادیر ارزش‌ها عملکرد **Regex entity extractor** را با عملکرد **CRF entity extractor** مقایسه کنید. (۱۵ نمره)

^۱ Test

^۲ Accuracy

^۳ F1 Score

^۴ Confusion Matrix