

هدف پروژه درس توسعه نرم‌افزار جدید و استفاده از ابزارهای پردازش زبان‌های طبیعی موجود به منظور اجرای تسک‌های از پیش تعریف شده پردازش زبان از قبیل *NER, Classification, POS Tagging, Parsing, Sentiment Analysis* و غیره می‌باشد. این پروژه در چند مرحله تعریف می‌شود. در مرحله اول به منظور ایجاد انگیزه و علاقه و همچنین تفاوت چالش‌ها بین پروژه‌های شما دانشجویان، لازم است شما مجموعه داده‌ای در ارتباط با موضوع مورد علاقه خود جمع‌آوری کنید. داشتن فرضیه‌ای در مورد داده‌های مورد علاقه خود که علاقمند باشید با پردازش آن، بتوانید آن فرضیه را تایید یا ابطال کنید، باعث جذاب‌تر شدن پروژه برای شما و دیگران خواهد بود.

هدف پروژه این درس تحقیق و نوآوری در روش‌ها و ابزارهای پردازش زبان‌های طبیعی نیست. هدف تسلط شما روی تسک‌ها و استفاده مناسب از آنها روی داده جدید می‌باشد. با توجه به داده جدید جمع‌آوری شده توسط شما، انتظار می‌رود که در نهایت هر دانشجو یک مقاله فنی در رابطه با نتایج به دست آمده در پروژه‌اش ارائه کند. البته چنانچه داده جدیدی جمع‌آوری و ارائه کنید ممکن است قابلیت چاپ مقاله علمی نیز داشته باشد.

چنانچه به تسک خاصی (مثلاً چت‌بات یا ترجمه ماشینی) علاقمند هستید، سعی کنید با راهنمایی استاد حل تمرین، داده خود را به گونه‌ای انتخاب کنید که علاوه بر تسک‌هایی که به تدریج اعلام می‌کنیم، مناسب تسک مورد علاقه شما نیز باشد. در بخش پایانی پروژه امکان انجام یک تسک انتخابی به جای تسک معرفی شده را خواهید داشت. برای آشنایی با تسک‌های مختلف می‌توانید [nlpprogress.com](http://nlpprogress.com) را ببینید.

مرحله اول پروژه انتخاب موضوع و جمع‌آوری داده برای آن می‌باشد. در این مرحله انتظار می‌رود شما به مجموعه داده جدید در رابطه با موضوع مورد علاقه خود جمع‌آوری کنید. لازم است که جمع‌آوری داده به گونه‌ای باشد که دارای دو دسته باشد که مناسب تحلیل مقایسه‌ای دو دسته و تسک طبقه‌بندی باشد. اندازه داده هر دسته باید حداقل ۵۰ هزار کلمه باشد.

مجموعه داده می‌تواند به هر زبانی باشد. ولی لازم است شما به آن موضوع آشنا و علاقمند بوده به نحوی که قبل از انجام تسک‌ها در مورد نتیجه تسک حدس و نظر داشته باشید و چنانچه نتیجه تحلیل با حدس شما متفاوت بود، بتوانید دیباگ کرده و علت این تفاوت را تحلیل کنید. برای شروع پروژه لازم است پیشنهاد پروژه‌ای در قالب یک فایل در گریداسکوپ ارسال نمایید. در پیشنهاد پروژه خود سعی کنید به سوال‌های زیر پاسخ دهید.

- موضوع داده
- زبان داده
- آیا داده به زبان محاوره‌ای است یا رسمی؟
- دسته‌بندی داده (حداقل ۲ دسته) به چه شکل است؟
- داده از چه جهت برای شما جذاب است؟
- آیا فرضیه‌ای در رابطه با داده دارید؟
- نحوه جمع‌آوری داده؟
- آیا برای جمع‌آوری داده از کتابخانه یا *api* خاصی استفاده خواهید کرد؟ آیا دسترسی به این کتابخانه *api* دارید؟
- چه محدودیتی برای جمع‌آوری داده دارید؟ پیش‌بینی می‌کنید چه حجم از داده بتوانید جمع‌آوری کنید؟

ایده اولیه پروپوزال را حتما در اسکرام هفتگی قبل از پایان سال با استاد حل تمرین خود قطعی کنید و پروپوزال را تا پایان سال ۱۳۹۹ در گرید اسکوپ ارسال نمایید. ارسال به موقع و کامل پیشنهاد پروژه شما ۵ درصد نمره پایانی پروژه را خواهد داشت.

موفق باشید

اعتمادی