

Information Retrieval

Doc A

Doc 1

Doc 2

Doc 3

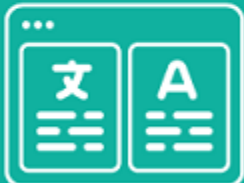
Sentiment Analysis



Information Extraction



Machine Translation



Question Answering



Human: When was Apollo sent to space?

Machine: First flight - AS-201, February 26, 1966

# پردازش متن در زبان فارسی با منابع محدود

Image courtesy of sigmoidal.io

یادگیری چند منظوره و همچنین استفاده از Active Learning در فرآیند برچسب گذاری میشوند.

**ارائه دهنده:** بهنام ثابتی در سال ۱۳۹۲ مدرک کارشناسی مهندسی کامپیوتر را از دانشگاه فردوسی مشهد و سپس در سال ۱۳۹۴ کارشناسی ارشد هوش مصنوعی را از دانشگاه صنعتی شریف اخذ کرده‌اند. ایشان از سال ۱۳۹۴ تا کنون دانشجوی دکتری هوش مصنوعی در دانشگاه صنعتی شریف هستند. حوزه‌ی تحقیقاتی ایشان یادگیری ماشین و شبکه‌های ژرف و استفاده آنها در پردازش زبان طبیعی است. ایشان از زمستان ۱۳۹۵ تاکنون در شرکت میراث به عنوان دانشمند داده و مدیر محصول مشغول به کار هستند.



دانشکده مهندسی کامپیوتر

شده است، که میتواند با منابع محدود نتایج قابل قبولی را تولید کنند: (۱) آموزش شبکه های عصبی برای حل مسائلی مانند مدلسازی زبان که دادگان حجیم برای آن در زبان فارسی وجود دارد، و استفاده از بازنمایی های تولید شده و Fine Tune کردن آنها جهت حل سایر مسائل پردازش متن. (۲) توسعهی سیاستهای Active Learning در راستای بهینه سازی فرآیند برچسب گذاری و تولید دادگان آموزشی. در این روش، داده هایی که بیشترین اطلاعات را در اختیار مدل قرار میدهند انتخاب شده و توسط کاربر برچسب گذاری میشوند، بدین ترتیب با کمترین هزینه ی ممکن میتوان داده های آموزشی با بیشترین بازدهی را تولید کرد. در این ارائه چالشهای تیم میراث در توسعهی سرویسهای پردازش زبان فارسی، مانند تحلیل تمایل، تحلیل احساس، دسته بندی متن، تشخیص موجودیتهای نامدار و تشخیص ناسزا، مورد بررسی قرار گرفته و رویکردهای استفاده شده در جهت حل این مسائل معرفی میشوند. این رویکردها شامل استفاده از مدلهای از پیش آموزش داده شده ی چند زبانه مانند BERT،

**سخنرانی های «داده»:**  
**پردازش متن در زبان فارسی با منابع محدود**

چهارشنبه ۲۴ مهر ۱۳۹۸  
ساعت ۱۴:۰۰ تا ۱۵:۳۰  
سالن خوارزمی (اتاق ۴۰۴)

**چکیده:** با توسعه ی روز افزون سرویسهای مبتنی بر وب و مشارکت کاربران در تولید محتوای متنی، داده های حجیم برای توسعه ی سامانه های مبتنی بر یادگیری عمیق در دسترس قرار گرفته است. هرچند دادگان و ابزارهای توسعه ی الگوریتمهای پردازش زبان طبیعی در زبانهایی مانند انگلیسی به وفور یافت میشود، اما پژوهشگران در زبان فارسی با کمبود منابع متنی برچسبگذاری شده مواجه اند. در راستای مقابله با این چالش روشهای متعددی پیشنهاد