

سری اول تمارین پردازش زبان طبیعی

محمدرضا اصولی - ۶۱۰۳۹۵۰۷۷

پاسخ سوال یک

پردازش زبان طبیعی زیرشاخه‌ای از زبان‌شناسی، علوم کامپیوتر، مهندسی اطلاعات و هوش مصنوعی با تمرکز بر ارتباط بین کامپیوتر و زبان طبیعی است. پردازش زبان طبیعی به طور کلی به برنامه‌ریزی کامپیوتر برای پردازش و آنالیز حجم زیادی داده زبان طبیعی می‌پردازد.

پردازش زبان طبیعی زیرمجموعه زبان‌شناسی محاسباتی است. در زبان‌شناسی محاسباتی به بررسی زبان و استفاده از کامپیوتر برای مدل‌سازی نحوه درک زبان توسط انسان می‌پردازیم، در صورتی پردازش زبان طبیعی بیشتر به ارتباط بین انسان و کامپیوتر از طریق زبان می‌پردازد.

از کاربردهای پردازش زبان طبیعی، می‌توان به سیستم‌های پاسخگویی هوشمند، طبقه‌بندی متون و پیش‌بینی در نوشتن متن‌ها اشاره کرد.

پاسخ سوال دو

۱. ایمیل:

$\text{\textasciitilde}([\text{\textbackslash}w_\\-\\.]^+)([\text{\textbackslash}w_\\-\\.]^+\\.[a-zA-Z]\{2,6\})\$$

که در آن علامت \wedge به معنای شروع رشته، علامت $\$$ به معنای پایان رشته، نماد w به معنای حروف بزرگ و کوچک انگلیسی به همراه ارقام، علامت $+$ به معنای وجود حداقل یک حرف می‌باشد.

۲. تاریخ با فرمت DD - MM - YYYY:

$\text{\textasciitilde}(1|2)\text{\textbackslash}d\{3\} \text{\textbackslash}- (0[1_9]|1[0_2]) \text{\textbackslash}- (0[1_9]|1\text{\textbackslash}d|2\text{\textbackslash}d|3[01])\$$

* به دلیل بهم‌ریختگی متن در Latex به جای - از _ استفاده شده است.

۳. آدرس URL

$\text{\textasciitilde}https?:\\\/\\\/(www\\.)?[-a-zA-Z0_9@: \%_\\+~\#=\{1,256\}\\.[a-zA-Z]\{2,6\}$
 $\text{\textbackslash}b([-a-zA-Z0_9()@: \%_\\+\\.~\#?&\\/=\]*)\$$

پاسخ سوال سه

به طور کلی در پردازش زبان طبیعی، تقسیم‌بندی (segmentation) به معنای تقسیم کردن متن به بخش‌های معنادار مثل کلمه، جمله یا مبحث است. این عمل هم در ذهن انسان و هم در پردازش‌های کامپیوتری وجود دارد. یکی از نشانه‌های اصلی برای پایان جملات علائم نگارشی مانند نقطه، علامت سوال و علامت تعجب است. اما موارد بسیاری وجود دارد که این علائم پایان جمله نیستند، در اینجا می‌توان یک طبقه‌بند دو حالتی برای تشخیص این موارد ساخت. به این صورت که در مورد هر نقطه در متن، این طبقه‌بند پایان جمله بودن یا نبودن را تشخیص می‌دهد. از روش‌های مختلفی برای ساخت این طبقه‌بند می‌توان استفاده کرد، مانند درخت‌های تصمیم یا گرامرهای دست‌ساز. همچنین از مدل‌های یادگیری ماشین نیز می‌توان برای ساخت این طبقه‌بند استفاده کرد.