پـــردازش زبـــان طبیعی با پایتون

مؤلفین

هابسون لین

کول هاوارد

هانس ماکس هاپکه

مترجم

ایوب ترکیان

نیاز دانش

فهرست مطالب

عنوان **شماره صفحه**

فصل ۱

مرور کلی NLP

شما در حال ورود به رویداد جالب پردازش زبان طبیعی (NLP) هستید. ابتدا، نشان داده خواهد شد NLP چیست و چه کارهایی با آن می‌توان انجام داد. این کار، چرخ‌ها را به حرکت در آورده، و به شما کمک می‌کند روش‌های استفاده از NLP در زندگی روزمره خویش، سر کار و در منزل، را بررسی کنید.

سپس، جزئیات دقیق نحوه پردازش مقدار کمی متن انگلیسی با استفاده از زبان برنامه‌سازی مثل پایتون بررسی شده، که کمک خواهد کرد به تدریج جعبه ابزار NLP خویش را درست کنید. در این فصل، اولین برنامه خویش را خواهید نوشت که می‌تواند جملات انگلیسی را خوانده و بنویسد. این قطعه برنامه پایتون یکی از تعداد زیاد برنامه‌هایی است که استفاده خواهید کرد، تا همه شگردهای مورد نیاز برای سرهم‌سازی یک موتور دیالوگ زبان انگلیسی، ماشین سخنگو، را یاد بگیرید.

۱.۱ زبان طبیعی و زبان برنامه‌سازی

زبان‌های طبیعی از زبان‌های برنامه‌سازی رایانه متفاوت هستند. قرار نیست آنها، مثل زبان‌های برنامه‌سازی، به یک مجموعه معین عملیات ریاضی، ترجمه شوند. زبان‌های طبیعی، مورد استفاده انسان‌ها برای به اشتراک‌گذاری اطلاعات با یکدیگر هستند. از زبان‌های برنامه‌سازی برای صحبت با یکدیگر راجع به وقایع روزانه یا برای آدرس مغازه خواروبارفروشی دادن، استفاده نمی‌شود. یک برنامه کامپیوتر نوشته شده با زبان برنامه‌سازی، به ماشین می‌گوید دقیقاً چه کاری را انجام بدهد. ولی برای زبان‌های طبیعی نظیر فارسی و عربی، کامپایلر یا مفسر وجود ندارد.

تعریف پردازش زبان طبیعی، یک زمینه تحقیق در علم کامپیوتر و هوش مصنوعی (AI) بوده که مربوط به پردازش زبان طبیعی نظیر انگلیسی یا ماندارین است. این پردازش به‌طور کلی شامل ترجمه زبان طبیعی به داده‌ها (اعداد) بوده که کامپیوتر بتواند برای یادگیری در مورد پدیده‌ها استفاده کند. و این شناخت جهان، در بعضی مواقع برای تولید متن زبان طبیعی استفاده شده، که انعکاس آن شناخت است.

با این حال، در این فصل نحوه پردازش زبان طبیعی توسط ماشین نشان داده می‌شود. ممکن است حتی این را یک مفسر زبان طبیعی، درست مشابه مفسر پایتون، در نظر بگیرید. در موقعی که برنامه کامپیوتری توسعه داده شده، زبان طبیعی را پردازش می‌کند، قادر به عمل روی آن عبارات یا حتی پاسخ به آنها خواهد بود. ولی این اقدامات و پاسخ‌ها دقیقاً تعریف نشده‌اند، که فضای اعمال سلیقه شما، توسعه‌گر خط مسیر زبان طبیعی، را فراهم می‌کند.

تعریف یک سیستم پردازش زبان طبیعی غالباً خط‌مسیر نامیده شده، چون معمولاً شامل چند مرحله پردازش بوده که در آنها زبان طبیعی از یک طرف وارد شده، و خروجی پردازش شده از طرف دیگر به بیرون می‌رود.

به زودی، توان نوشتن نرم‌افزاری را در خود ایجاد خواهید کرد که کارهای جالب غیرقابل پیش‌بینی، نظیر انجام گفتگو را طوری صورت داده که به نظر برسد ماشین‌ها یک مقدار رفتار انسانی دارند. در ابتدا، ممکن است کاری که تکنولوژی پیشرفته انجام می‌دهد، یک مقدار سحرآمیز جلوه کند. ولی پرده که عقب زده شود، پشت صحنه را می‌توان بررسی کرد، و زود همه ابزار مورد نیاز برای انجام شگردهای جادویی را خود شما کشف خواهید کرد.

۲.۱ جادو

چه چیزی در باره اینکه یک ماشین بتواند در یک زبان طبیعی بخواند و بنویسد، جادویی است؟ ماشین‌ها از زمان اختراع کامپیوتر، زبان‌ها را پردازش میکرده‌اند. با این حال، این زبان‌های «رسمی»- نظیر زبان‌های اولیه Ada، COBOL، و فورترن، طوری طراحی می‌شدند که فقط به یک شیوه صحیح تفسیر (یا کامپایل) شوند. امروزه، ویکی‌پدیا لیستی از بیش از ۷۰۰ زبان برنامه‌سازی را فراهم می‌کند. بر خلاف این، ۱۰ برابر این مقدار زبان طبیعی محاوره‌ای در جهان وجود دارد. نمایه گوگل اسناد زبان طبیعی بیش از ۱۰۰ میلیون گیگابایت است. و این فقط نمایه است. و ناکامل است. اندازه محتوای زبان طبیعی واقعی برخط، در حال حاضر متجاوز از ۱۰۰