

پردازش زبان طبیعی با پایتون

نویسندگان:

استیون برد، اوان کلین، ادوارد لوپر

مترجمان:

علی ناصراسدی

عضو هیأت علمی گروه کامپیوتر، مجتمع آموزش عالی زنجان

علی رهنما

عضو هیأت علمی گروه کامپیوتر، مجتمع آموزش عالی زنجان

سرشناسه	: برد، استیون
عنوان و نام پدیدآور	: پردازش زبان طبیعی با پایتون/ نویسندگان استیون برد، اوان کلین، ادوارد لوپر ! مترجمان علی ناصر اسدی، علی رهنما.
مشخصات نشر	: تهران: آرنا، ۱۳۹۸.
مشخصات ظاهری	: ۵۸۴ ص.
شابک	: ۹۷۸-۶۰۰-۳۵۶۸۷۴-۷
وضعیت فهرست نویسی	: فیبا
پادداشت	: عنوان اصلی: Natural Language Processing with Python.
موضوع	: پردازش زبان طبیعی
موضوع	: (Natural language processing (Computer science
موضوع	: پایتون (زبان برنامه نویسی کامپیوتر)
موضوع	: (Python (Computer program language
شناسه افزوده	: کلاین، یوان
شناسه افزوده	: Klein, Ewan
شناسه افزوده	: لوپر، ادوارد
شناسه افزوده	: Klein, Edward
شناسه افزوده	: رهنما، علی، ۱۳۵۹-، مترجم
شناسه افزوده	: ناصر اسدی، علی، ۱۳۶۱-، مترجم
رده بندی کنگره	: ۵۰۷۶/۹
رده بندی دیویی	: ۰۰۴/۳۵
شماره کتابشناسی ملی	: ۱۰۳۱



نشر آرنا

عنوان	: پردازش زبان طبیعی با پایتون
نویسندگان	: استیون برد، اوان کلین، ادوارد لوپر
مترجمان	: علی ناصر اسدی- علی رهنما
ناشر	: آرنا
چاپ	: اول ۱۳۹۸
شمارگان	: ۵۰۰ نسخه قیمت: ۱۱۰۰۰۰ تومان
شابک	: ۹۷۸-۶۰۰-۳۵۶-۸۷۴-۷

www.arnapub.com

فهرست مطالب

۱۱	پیشگفتار متنی
۱۳	پیشگفتار

۲۵	فصل ۱- پردازش زبان و پایتون
۲۶	۱-۱ محاسبات با زبان: متن‌ها و کلمات
۳۷	۲-۱ نگاهی دقیق‌تر به پایتون: متغیرها و لیست‌ها
۴۶	۳-۱ محاسبه با زبان: آمار ساده
۵۳	۴-۱ بازگشت به پایتون: تصمیم‌گیری و کنده‌کاری
۶۰	۵-۱ درک خودکار زبان طبیعی
۶۸	۶-۱ خلاصه
۶۹	۷-۱ مطالعه‌ی بیشتر
۷۰	۸-۱ تمرین‌ها

۷۳	فصل ۲- دسترسی به پیکره‌های متنی و منابع لغوی
۷۳	۱-۲ دسترسی به پیکره‌های متنی
۸۹	۲-۲ توزیع فراوانی شرطی

۹۵	کار بیشتر با پایتون: استفاده مجدد از کدها	۳-۲
۹۹	منابع لغوی	۴-۲
۱۰۹	WordNet	۵-۲
۱۱۷	خلاصه	۶-۲
۱۱۷	مطالعه‌ی بیشتر	۷-۲
۱۱۸	تمرین‌ها	۸-۲

فصل ۳- پردازش متن خام

۱۲۴	استرس: متن از طریق وب یا دیسک	۱-۳
۱۳۳	رشته‌ها: بخشش متن در پایین سطح	۲-۳
۱۴۲	پردازش متن: یونیکد	۳-۳
۱۴۶	عبارات باقاعده برای تشخیص الگوهای کلمه	۴-۳
۱۵۲	کاربردهای مفید عبارت باقاعده	۵-۳
۱۵۹	نرمال‌سازی متن	۶-۳
۱۶۲	عبارات باقاعده برای تجزیه‌ی لغوی متن	۷-۳
۱۶۶	بخش‌بندی	۸-۳
۱۷۱	قالب‌بندی: از لیست‌ها تا رشته‌ها	۹-۳
۱۷۸	خلاصه	۱۰-۳
۱۷۹	مطالعه‌ی بیشتر	۱۱-۳
۱۸۱	تمرین‌ها	۱۲-۳

فصل ۴- نوشتن برنامه‌های ساخت یافته

۱۹۰	برگشت به مفاهیم اصلی	۱-۴
۱۹۴	دنباله‌ها	۲-۴

۲۰۱	مباحثی در مورد سبک برنامه‌نویسی	۳-۴
۲۰۶	توابع: اساس برنامه‌نویسی ساخت یافته	۴-۴
۲۱۵	کار بیشتر با توابع	۵-۴
۲۲۲	توسعه‌ی برنامه	۶-۴
۲۳۰	طراحی الگوریتم	۷-۴
۲۳۸	نمونه‌ای از کتابخانه‌های پایتون	۸-۴
۲۴۴	خلاصه	۹-۴
۲۴۴	مطالعه‌ی بیشتر	۱۰-۴
۲۴۵	تمرین‌ها	۱۱-۴

فصل ۵- رده‌های زبرچسب‌زنی کلمات

۲۵۱	استفاده از یک زبرچسب‌زن	۱-۵
۲۵۴	پیکره‌های زبرچسب خورده	۲-۵
۲۶۴	نگاشت کلمه‌ها به ویژگی‌های استفاده از واژه‌نامه‌های پایتون	۳-۵
۲۷۷	زبرچسب‌زنی خودکار	۴-۵
۲۸۲	زبرچسب زدن چندنگاشتی‌ها	۵-۵
۲۹۰	زبرچسب‌زنی مبتنی بر تغییر شکل	۶-۵
۲۹۲	چگونگی تعیین رده‌ی یک کلمه	۷-۵
۲۹۵	خلاصه	۸-۵
۲۹۶	مطالعه‌ی بیشتر	۹-۵
۲۹۸	تمرین‌ها	۱۰-۵

فصل ۶- یادگیری دسته‌بندی متن

۳۰۵	دسته‌بندی با نظارت	۱-۶
-----	--------------------	-----

۳۲۱	۶-۲	مثال‌های بیشتری از دسته‌بندی با نظارت
۳۲۶	۶-۳	ارزیابی
۳۳۲	۶-۴	درخت‌های تصمیم
۳۳۷	۶-۵	دسته‌بندهای نایوبیز
۳۴۳	۶-۶	دسته‌بندهای آنروپی بیشینه
۳۴۸	۶-۷	مدل‌سازی الگوهای زبان‌شناسی
۳۴۹	۶-۸	خلاصه
۳۵۰	۶-۹	مطالعه بیشتر
۳۵۲	۶-۱۰	تمرین‌ها

فصل ۷- استخراج اطلاعات از متن

۳۵۵	۷-۱	استخراج اطلاعات
۳۵۹	۷-۲	قطعه‌بندی
۳۶۷	۷-۳	توسعه و ارزیابی قطعه‌بندها
۳۷۵	۷-۴	بازگشت در ساختار زبان‌شناسی
۳۸۰	۷-۵	تشخیص موجودیت‌های نامدار
۳۸۳	۷-۶	استخراج رابطه
۳۸۵	۷-۷	خلاصه
۳۸۶	۷-۸	مطالعه‌ی بیشتر
۳۸۷	۷-۹	تمرین‌ها

فصل ۸- تحلیل ساختار جمله

۳۹۱	۸-۱	برخی مشکلات گرامری
۳۹۶	۸-۲	کاربردهای نحو کدام‌اند؟

۳۹۹	گرامر مستقل از متن	۳-۸
۴۰۴	تجزیه با گرامر مستقل از متن	۴-۸
۴۱۴	وابستگی‌ها و گرامر وابستگی	۵-۸
۴۲۰	توسعه‌ی گرامر	۶-۸
۴۲۷	خلاصه	۷-۸
۴۲۸	مطالعه‌ی بیشتر	۸-۸
۴۲۹	تمرین‌ها	۹-۸

فصل ۹- ساختن گرامرهای مبتنی بر ویژگی

۴۳۵	انزگه‌ی گرامری	۱-۹
۴۴۷	پردازش ساختارهای ویژگی	۲-۹
۴۵۵	توسعه‌ی یک گرامر مبتنی بر ویژگی	۳-۹
۴۶۷	خلاصه	۴-۹
۴۶۸	مطالعه‌ی بیشتر	۵-۹
۴۷۰	تمرین‌ها	۶-۹

فصل ۱۰- تحلیل معنی جملات

۴۷۳	درک زبان طبیعی	۱-۱۰
۴۷۳	منطق گزاره‌ای	۲-۱۰
۴۸۱	منطق مرتبه‌ی اول	۳-۱۰
۴۸۶	معناشناسی جملات انگلیسی	۴-۱۰
۵۰۲	معناشناسی گفتمان	۵-۱۰
۵۱۶	خلاصه	۶-۱۰
۵۲۴	مطالعه‌ی بیشتر	۷-۱۰
۵۲۵		

۵۲۶ ۸-۱۰ تمرین‌ها

فصل ۱۱- مدیریت داده‌های زبان‌شناسی ۵۲۹

۵۲۹ ۱-۱۱ ساختار پیکره: یک مطالعه‌ی موردی

۵۳۴ ۲-۱۱ چرخه‌ی حیات یک پیکره

۵۴۰ ۳-۱۱ جمع‌آوری داده‌ها

۵۵۱ ۴-۱۱ کار با XML

۵۶۰ ۵-۱۱ کار با داده‌های Toolbox

۵۶۴ ۶-۱۱ توصیف منابع زبانی با استفاده از فراداده‌های OLAC

۵۶۷ ۷-۱۱ خلاصه

۵۶۷ ۸-۱۱ مطالعه‌ی بیشتر

۵۶۹ ۹-۱۱ تمرین‌ها

فصل ۱۲- سخن پایانی: چالش زبان ۵۷۱

۵۷۲ ۱-۱۲ پردازش زبان در مقابل پردازش نماد

۵۷۴ ۲-۱۲ شکاف‌های فلسفی معاصر

۵۷۵ ۳-۱۲ نقشه‌ی راه پردازش زبان طبیعی

۵۷۷ ۴-۱۲ کلام آخر

۵۷۹ منابع و مراجع

پیشگفتار مترجمان

این مجله زبان یکی از مهم‌ترین ابزارهای بشر برای برقراری رابطه با دیگران و سعی در بیان و انتشار عقاید و خواسته‌های فردی و اجتماعی بوده است. از سوی دیگر، نقش مهم زبان در گسترش و انتشار فرهنگ‌ها، مکاتب فکر و حتی میل به مقاصد سیاسی، اجتماعی و اقتصادی، انکارناپذیر است. به همین دلیل، هم‌زمان با اولین برقه‌ها، علم کامپیوتر و یکی از زیرشاخه‌های جذاب و پرکاربرد آن یعنی هوش مصنوعی، در ذهن بشر (دوره ۱۹۰۰ میلادی و به‌واسطه تلاش‌های دانشمند بزرگی مانند آلن تورینگ)، محققان به سمت استفاده از کامپیوتر به‌منظور درک زبان طبیعی و توانا ساختن ماشین‌ها به برقراری ارتباط با انسان از طریق زبان طبیعی تاایل پیدا کرده‌اند. از دیدگاه تورینگ، ماشینی هوشمند است که توانایی درک و تولید زبان طبیعی انسان و آماده‌سازی آن برای تعامل با انسان‌ها را داشته باشد. به غیر از این هدف، کاربردهای دیگری نیز به‌مرو برای این توانایی مطرح شد که از مهم‌ترین آن‌ها می‌توان به ترجمه ماشینی، خلاصه‌سازی متن‌های زبان طبیعی، پاسخگویی به سؤالات مطرح‌شده به زبان طبیعی، بازیابی و استخراج اطلاعات از متن‌های غیر ساخت‌یافته و غیره اشاره کرد.

به‌طور کلی به فرایند درک و تولید زبان طبیعی با استفاده از ابزارهای حساس، **پردازش زبان طبیعی**، یا مطابق نظر برخی از متخصصان و افراد خبره، **زبان‌شناسی محاسبه‌ای** گفته می‌شود. با این حال، علی‌رغم پیشرفت چشمگیری که در نیم‌قرن گذشته و به‌ویژه در چند دهه‌ی اخیر در زمینه‌ی علوم کامپیوتر و هوش مصنوعی رخ داده است، هنوز پدیده‌ی پردازش زبان طبیعی با میزان دقت و کیفی که ایده آل محققان و سیستم‌های اطلاعاتی و ارتباطی آرمانی است، فاصله‌ی بسیاری دارد. از مهم‌ترین عوامل تأثیرگذار بر این مشکل می‌توان به مواردی مانند وجود ابهام (از نوع نحوی، ابهام ناشی از معنای متعدد کلمات یا ابهام مربوط به حساسیت معنی جمله به زمینه‌ی مورد استفاده) در زبان طبیعی، نامشخص بوده ساختار نحوی بسیاری از زبان‌های طبیعی اشاره کرد به گونه‌ای که از دیدگاه بسیاری از صاحب‌نظران، این مسئله یک مسئله‌ی کامل در هوش مصنوعی به‌حساب می‌آید.