

۲۰۵۲۹۳۰

پردازش زبان طبیعی با پایتون

مؤلفین

هابسون لین

کول هاوارد

هانس ماکس هاپکه

مترجم

ایوب ترکیان

نیاز دانش



سرشناسه	لین، هوبسن	Lane, Hobson
عنوان و نام پدیدآور	پردازش زبان طبیعی با پایتون / مولفین هابسون لین، کول هاوارد، هانس ماکس هاپکه	
مشخصات نشر	تهران: نیاز دانش، ۱۳۹۸.	
مشخصات ظاهری	۵۲۴ ص.: مصور.	
شابک	978-600-8906-54-4	
وضعیت فهرست‌نویسی	فیا	
یادداشت	عنوان اصلی: Natural language processing in action: understanding, analyzing, and generating text with Python., [2019]	
موضوع	پردازش زبان طبیعی	Natural language processing (Computer science)
موضوع	پایتون (زبان برنامه‌نویسی کامپیوتر)	Python (Computer program language)
شناسه افزوده	هاوارد، کول	Howard, Cole
شناسه افزوده	هاپکه، هانس ماکس	Hapke, Hannes Max
شناسه افزوده	ترکیان، ایوب، ۱۳۳۷، مترجم.	
رده‌بندی دنگ	Q۸۷۶/۹	
رده‌بندی دیویی	۰۰۶/۳۸	
شماره کتابشناسی	۵۸۳۵۱۰۸	



نیاز دانش

نام کتاب	پردازش زبان طبیعی با پایتون
مؤلفین	هابسون لین - کرن هاوارد - هانس ماکس هاپکه
مترجم	ایوب ترکیان
مدیر اجرایی - ناظر بر چاپ	حمیدرضا محمد شیرازی - محمد شمس
ناشر	نیاز دانش
صفحه‌آرا	واحد تولید انتشارات نیاز دانش
نوبت چاپ	اول - ۱۳۹۸
شمارگان	۱۰۰ نسخه
قیمت	۸۴۰۰۰۰ ریال

ISBN:978-600-8906-54-4

شابک: ۹۷۸-۶۰۰-۸۹۰۶-۵۴-۴

هرگونه چاپ و تکثیر (اعم از زیراکس، بازنویسی، ضبط کامپیوتری و تهیه‌ی CD) از محتویات این اثر بدون اجازه‌ی کتبی ناشر ممنوع است، متخلفان به موجب بند ۵ از ماده‌ی ۲ قانون حمایت از مؤلفان، مصنفان و هنرمندان تحت پیگرد قانونی قرار می‌گیرند.

کلیه حقوق این اثر برای ناشر محفوظ است.

آدرس انتشارات: تهران، میدان انقلاب، خیابان ۱۲ فروردین، تقاطع وحید نظری، پلاک ۲۵۵، طبقه ۱، واحد ۲

۰۲۱-۶۶۴۴۷۸۱۰۶-۶۶۴۴۷۸۱۰۸-۰۹۱۲۷۰۷۳۹۳۵

www.Niaze-Danesh.com

مشاوره جهت نشر: ۰۹۱۲-۲۱۰۶۷۰۹

فهرست مطالب

شماره صفحه

عنوان

۹	فصل ۱ / مرور کلی NLP
۹	۱.۱ زبان طبیعی و زبان برنامه‌سازی
۱۰	۲.۱ جادو
۱۱	۱.۲.۱ ماشین سخنگو
۱۲	۲.۲.۱ ریاضی
۱۴	۳.۱ کاربردهای عملی
۱۶	۴.۱ زبان: نگاه رایانه
۱۷	۱.۴.۱ زبان قضاها
۱۸	۴.۱ عبارت منظم
۲۰	۳.۴.۱ روایات سخنگوی ساده
۲۴	۱.۴.۱ روش دیگر
۲۹	۵.۱ فراقضا
۳۱	۶.۱ ترتیب واژه و اراسر
۳۲	۷.۱ خط‌مسیر زبان طبیعی روی گراف
۳۵	۸.۱ پردازش در عمق (عمیق)
۳۷	۹.۱ IQ زبان طبیعی
۴۰	۱۰.۱ خلاصه

۴۱	فصل ۲ / توکن‌سازی واژه
۴۳	۱.۲ چالش‌ها (پیش‌نمایش ریشه‌یابی)
۴۴	۲.۲ ساخت فرهنگ لغات با توکن‌ساز
۵۴	۱.۲.۲ ضرب داخلی
۵۵	۲.۲.۲ سنجش هم‌پوشانی کیف‌واژه
۵۶	۳.۲.۲ بهبود توکن
۶۲	۴.۲.۲ بسط فرهنگ‌نامه با n-گرام‌ها
۶۸	۵.۲.۲ نرمال‌سازی قاموس
۷۶	۳.۲ تمایل
۷۹	۱.۳.۲ VADER - تحلیل‌گر تمایل قاعده‌محور
۸۰	۲.۳.۲ بیز ساده‌انگارانه
۸۴	۴.۲ خلاصه

۸۵	فصل ۳ / ریاضی کلمات (بردارهای TF-IDF)
۸۶	۱.۳ کیف‌واژه
۹۱	۲.۳ بردارسازی
۹۴	۱.۲.۳ فضاها و برداری
۹۹	۳.۳ قانون زیف
۱۰۲	۴.۳ مدل‌سازی موضوعی
۱۰۵	۱.۴.۳ بازگشت به زیف
۱۰۶	۲.۴.۳ درجه‌بندی مربوط بودن

۱۰۹	ابزار	۳.۴.۳
۱۱۰	التراتیوها	۴.۴.۳
۱۱۰	okapi BM25	۵.۴.۳
۱۱۲	مرحله بعدی	۶.۴.۳
۱۱۲	خلاصه	۵.۳

فصل ۴ / تحلیل معنایی

۱۱۳	۱.۴	تعداد کلمه و نمرات موضوع
۱۱۴	۱.۱.۴	بردارهای TF-IDF و لم سازی
۱۱۴	۲.۱.۴	بردارهای موضوعی
۱۱۶	۳.۱.۴	آزمایش ذهنی
۱۱۸	۴.۱.۴	الگوریتم رتبه بندی موضوعی
۱۲۲	۵.۱.۴	طبقه گر LDA
۱۲۴	۲.۴	تجزیه سمانتیک نهفته
۱۲۹	۱.۲.	رمایش ذهنی واقعی
۱۳۲	۳.۴	تجزیه مقدار ویژه
۱۳۴	۱.۳.۴	U - بردار وینچ
۱۳۶	۲.۳.۴	S - مقادیر وینچ
۱۳۷	۳.۳.۴	VT بردارهای وینچ است
۱۳۸	۴.۳.۴	جهت مارتینس S
۱۳۹	۵.۳.۴	کوتاه سازی موسوعه
۱۴۰	۴.۴	آنالیز مؤلفه اصلی
۱۴۲	۱.۴.۴	PCA روی بردارهای - بعدی
۱۴۴	۲.۴.۴	بازگشت به NLP
۱۴۵	۳.۴.۴	استفاده PCA در تحلیل معنایی S
۱۴۸	۴.۴.۴	SVD کوتاه شده تحلیل معنایی
۱۵۰	۵.۴.۴	دسته بندی اسیم با LSA
۱۵۱	۵.۴	تخصیص دیریکله نهفته (LDiA)
۱۵۴	۱.۵.۴	ایده LDiA
۱۵۵	۲.۵.۴	مدل موضوعی LDiA پیام کوتاه
۱۵۷	۳.۵.۴	$LDA + LDiA$ = طبقه گر اسیم
۱۶۰	۴.۵.۴	مقایسه متصفانه: ۳۲ موضوع LDiA
۱۶۲	۶.۴	فاصله و مشابهت
۱۶۴	۷.۴	حرکت با پس خور
۱۶۷	۱.۷.۴	تحلیل تمایز خطی
۱۶۸	۸.۴	قدرت بردار موضوعی
۱۷۰	۱.۸.۴	جستجوی سمانتیک
۱۷۰	۲.۸.۴	بهبودها
۱۷۳	۹.۴	خلاصه
۱۷۳		

فصل ۵ / مبانی شبکه های عصبی

۱۷۵	۱.۵	شبکه های عصبی
۱۷۶	۱.۱.۵	پرسپترون
۱۷۶	۲.۱.۵	پرسپترون عددی
۱۷۷	۳.۱.۵	انحرافی به پایاس
۱۷۸		

۱۹۳	۴.۱.۵	سطح خطا
۱۹۴	۵.۱.۵	شیب
۱۹۵	۶.۱.۵	اصلاح رویه
۱۹۷	۷.۱.۵	کراس: شبکه‌های عصبی در پایتون
۲۰۰	۸.۱.۵	شبکه‌های عمیق
۲۰۰	۹.۱.۵	نرمال‌سازی
۲۰۱	۲.۵	خلاصه

۲۰۳	فصل ۶ / استدلال با بردارهای واژه	
۲۰۴	۱۶	پرسمان‌های سمانتیک و تشبیه
۲۰۴	۱.۱۶	سوالات تشبیهی
۲۰۶	۲۶	برداری واژه
۲۱۰	۳۶	استدلال با گرایش برداری
۲۱۳	۲.۱	حاسبه نمایش‌های Word2vec
۲۲۳	۲۶	مازول genism.word2vec
۲۲۵	۴.۲۶	رلید نیس برای واژه
۲۲۸	۵.۲۶	GloV و Word2vec
۲۲۹	۶.۲۶	FastText
۲۲۹	۷.۲۶	FastText و Word2vec
۲۳۱	۸.۲۶	مصورسازی روابط واژه
۲۳۸	۹.۲۶	واژه‌های غیرطبیعی
۲۳۸	۱۰.۲۶	مشابهت سند با Doc2vec
۲۴۱	۳.۶	خلاصه

۲۴۴	فصل ۷ / شبکه عصبی کانولوشن	
۲۴۴	۱.۷	یادگیری معنا
۲۴۶	۲.۷	چیمابزار
۲۴۷	۳.۷	شبکه عصبی کانولوشن
۲۴۷	۱.۳.۷	بلوک‌های ساخت
۲۵۰	۲.۳.۷	اندازه گام (پرش)
۲۵۰	۳.۳.۷	ترکیب فیلتر
۲۵۲	۴.۳.۷	لایه گذاری
۲۵۴	۵.۳.۷	یادگیری
۲۵۵	۴.۷	پنجره‌های کم‌عرض
۲۵۶	۱.۴.۷	بیاده‌سازی در کراس: آماده‌سازی داده‌ها
۲۶۲	۲.۴.۷	معماری شبکه عصبی کانولوشن
۲۶۳	۳.۴.۷	رای گیری
۲۶۶	۴.۴.۷	دورریزی
۲۶۷	۵.۴.۷	گیلاس روی بستنی
۲۶۹	۶.۴.۷	یادگیری (آموزش)
۲۷۱	۷.۴.۷	استفاده از مدل در خط‌میسر
۲۷۳	۸.۴.۷	گام بعدی
۲۷۴	۵.۷	خلاصه

۲۷۵	فصل ۱۸ / شبکه عصبی برگشتی
۲۷۸	۱۸ خاطرآوری با شبکه‌های برگشتی
۲۸۳	۱.۱۸ پس‌انتشار در زمان
۲۸۶	۲.۱۸ به‌روزرسانی
۲۸۸	۳.۱۸ خلاصه
۲۸۹	۴.۱۸ هزینه محاسباتی
۲۸۹	۵.۱۸ شبکه عصبی برگشتی با کراس
۲۹۴	۲۸ سر هم کردن موارد
۲۹۶	۲۸ یادگیری
۲۹۷	۴۸ فرایارامترها
۳۰۰	۵۸ پیش‌بینی کردن
۳۰۱	۱.۵۸ حالت‌دار بودن
۳۰۲	۲.۵۸ جاده دوطرفه
۳۰۴	۳.۵۸ این چیست؟
۳۰۴	۶۸ خلاصه

۳۰۵	فصل ۹ / شبکه عصبی LSTM
۳۰۶	۱.۹ LSTM
۳۱۶	۱.۱.۹ پس‌انتشار در
۳۱۹	۲.۱.۹ پیش‌بینی
۳۲۰	۳.۱.۹ داده‌های غیرتکراری
۳۲۲	۴.۱.۹ بازگشت به داده‌های تکراری
۳۲۵	۵.۱.۹ کلمات و حروف
۳۳۱	۶.۱.۹ چت کردن
۳۳۴	۷.۱.۹ گفتار واضح
۳۴۱	۸.۱.۹ محتوای گفتار
۳۴۲	۹.۱.۹ انواع دیگر حافظه
۳۴۲	۱۰.۱.۹ تعمیق
۳۴۴	۲.۹ خلاصه

۳۴۵	فصل ۱۰ / مدل‌های توالی به توالی و توجه
۳۴۵	۱.۱۰ معماری رمزگذار-رمزدا
۳۴۷	۱.۱.۱۰ رمزگذاری فکر
۳۵۰	۲.۱.۱۰ مشابهت با خودرمزگذار
۳۵۱	۳.۱.۱۰ گفتگوی توالی به توالی
۳۵۱	۴.۱.۱۰ مرور LSTM
۳۵۲	۵.۱۰ سرهم‌سازی خط‌مسیر توالی به توالی
۳۵۳	۶.۱.۱۰ آماده‌سازی برای آموزش
۳۵۴	۷.۲.۱۰ مدل توالی به توالی در کراس
۳۵۵	۸.۲.۱۰ رمزگذار توالی
۳۵۶	۹.۲.۱۰ رمزگذاری فکر
۳۵۸	۱۰.۲.۱۰ سرهم‌سازی شبکه توالی به توالی
۳۵۹	۱۱.۱۰ آموزش شبکه توالی به توالی
۳۶۰	۱۲.۱۰ تولید توالی‌های خروجی

۳۶۱		۴.۱.۰	ساخت بات چت
۳۶۲		۱.۴.۱.۰	آماده‌سازی واژگان برای آموزش
۳۶۳		۲.۴.۱.۰	ساخت واژه‌نامه کاراکتر
۳۶۴		۳.۴.۱.۰	تولید مجموعه‌های آموزش تک‌نمادی
۳۶۴		۴.۴.۱.۰	آموزش بات چت
۳۶۵		۵.۴.۱.۰	سرهم‌سازی مدل برای تولید توالی
۳۶۵		۶.۴.۱.۰	پیش‌بینی توالی
۳۶۶		۷.۴.۱.۰	تولید پاسخ
۳۶۷		۸.۴.۱.۰	گفتگو با بات چت
۳۶۷		۵.۱.۰	بهبودها
۳۶۸		۱.۵.۱.۰	کاهش پیچیدگی آموزش
۳۶۹		۲.۵.۱.۰	توجه کردن
۳۷۰		۶.۱.۰	در دنیای واقعی
۳۷۲		۷.۱.۰	خلاصه

۳۷۳		فصل ۱۱ / استخراج اطلاعات	
۳۷۳		۱.۱.۱	هستارهای نام‌دار و روابط
۳۷۴		۱.۱.۱.۱	پایگاه دانش
۳۷۷		۲.۱.۱.۱	استخراج اطلاعات
۳۷۸		۲.۱.۱	الگوهای منظم
۳۷۹		۱.۲.۱.۱	عبارات منظم
۳۸۰		۲.۲.۱.۱	استخراج ویژگی ML
۳۸۲		۳.۱.۱	اطلاعات با ارزش استخراج
۳۸۲		۱.۳.۱.۱	استخراج محل‌های GIS
۳۸۳		۲.۳.۱.۱	استخراج تاریخ
۳۸۸		۴.۱.۱	استخراج ارتباطات (روابط)
۳۸۹		۱.۴.۱.۱	برچسب‌زنی اذات سخن (POS)
۳۹۳		۲.۴.۱.۱	نرمال‌سازی هستار نام‌دار
۳۹۴		۳.۴.۱.۱	نرمال‌سازی و استخراج ارتباط
۳۹۵		۴.۴.۱.۱	الگوهای کلمه
۳۹۵		۵.۴.۱.۱	بخش‌بندی
۳۹۷		۶.۴.۱.۱	کار نکردن تقسیم («؟!»)
۳۹۸		۷.۴.۱.۱	بخش‌بندی جمله با عبارت منظم
۴۰۰		۵.۱.۱	در دنیای واقعی
۴۰۱		۶.۱.۱	خلاصه

۴۰۳		فصل ۱۲ / موتورهای دیالوگ	
۴۰۴		۱.۱۲	مهارت زبان
۴۰۵		۱.۱.۱۲	رویکردهای جدید
۴۱۱		۲.۱.۱۲	رویکرد هیبریدی
۴۱۱		۲.۱۲	رویکرد انطباق الگو
۴۱۳		۱.۲.۱۲	بات چت انطباق الگو با AIML
۴۲۱		۲.۲.۱۲	نمای شبکه‌ای انطباق الگو
۴۲۱		۳.۱۲	استقرار
۴۲۳		۴.۱۲	بازیابی (جستجو)

۴۲۲	چالش سیاق	۱.۴.۱۲
۴۲۶	نمونه بات چت بازیابی بایه	۲.۴.۱۲
۴۳۰	بات چت جستجو بایه	۳.۴.۱۲
۴۳۳	مدل های زایشی (مولد)	۵.۱۲
۴۳۳	چت در مورد nlpia	۱.۵.۱۲
۴۳۵	محاسن و معایب دو رویکرد	۲.۵.۱۲
۴۳۵	ماشین قدرتی	۶.۱۲
۴۳۶	Will	۱۶.۱۲
۴۳۷	فرایند طراحی	۷.۱۲
۴۴۰	شگرد	۸.۱۲
۴۴۰	سوالات با پاسخ قابل پیش بینی	۱.۸.۱۲
۴۴۱	جذاب بودن	۲.۸.۱۲
۴۴۲	جستجو	۳.۸.۱۲
۴۴۲	اقبال	۴.۸.۱۲
۴۴۲	وصل یا فصل	۵.۸.۱۲
۴۴۲	احساسی شدن	۶.۸.۱۲
۴۴۳	د. بیای واقعی	۱۲
۴۴۴	خلاصه	۱۰.۱۲

۴۴۵	فصل ۱۳ / فرایند یادگیری	
۴۴۵	داده ها	۱.۱۳
۴۴۶	بهینه سازی الگوریتم های NLP	۲.۱۳
۴۴۷	نمایه سازی	۱.۲.۱۳
۴۴۸	نمایه سازی پیشرفته	۲.۲.۱۳
۴۵۰	نمایه سازی پیشرفته با Annoy	۳.۲.۱۳
۴۵۵	کاربرد نمایه های تقریبی	۴.۲.۱۳
۴۵۶	نمایه های تقریبی	۵.۲.۱۳
۴۵۷	الگوریتم های RAM ثابت	۳.۱۳
۴۵۷	Gensim	۱.۳.۱۳
۴۵۸	محاسبه گراف	۲.۳.۱۳
۴۵۹	موازی سازی محاسبات NLP	۴.۱۳
۴۵۹	آموزش مدل های NLP روی GPU	۱.۴.۱۳
۴۶۱	اجاره یا خرید	۲.۴.۱۳
۴۶۱	گزینه های اجاره GPU	۳.۴.۱۳
۴۶۲	واحدهای پردازش تانسور	۴.۴.۱۳
۴۶۲	کاهش جابجایی حافظه در زمان آموزش مدل	۵.۱۳
۴۶۶	اشراف به مدل با TensorBoard	۶.۱۳
۴۶۶	مصور سازی نهفته سازی کلمه	۱۶.۱۳
۴۶۸	خلاصه	۷.۱۳

۴۶۹	پیوست الف/ ابزار NLP	
۴۷۷	پیوست ب/ پایتون سرگرمی و عبارات منظم	
۴۸۳	پیوست ج/ بردارها و ماتریس ها	
۴۸۹	پیوست د/ ابزارها و تکنیک های یادگیری ماشین	
۵۰۳	پیوست ه/ برپایش AWS GPU	
۵۱۷	پیوست و/ هش سازی حساس به محیط	