به نام خدا

وب کاوی با

Scrapy,Selenium,Beautiful soup

مؤلفان:

نیما شفیعی رضوانی نژاد

افشین اسمعیل زاد آهندانی

مؤسسه فرهنگی هنری
دیباگران تهران

عنوان کتاب: وب کاوی با
Scrapy, Selenium, Beautiful soup

مولفان : نیما شفیعی رضوانی نژاد- افشین اسعمیل زاد آهندانی

ناشر: موسسه فرهنگی هنری دیباگران تهران
ویراستار: مهدیه مخبری
صفحه آرایی:فرنوش عبدالهی
طراح جلد:محمد اسمعیل زادآهندانی
نوبت چاپ: اول
تاریخ نشر: ۱۴۰۱
چاپ و صحافی:صدف
تیراژ:۱۰۰ جلد
قیمت:۱۲۰۰۰۰۰ ریال
شابک: ۹۷۸-۶۲۲-۲۱۸-۶۴۹-۴

نشانی واحد فروش:تهران، خیابان انقلاب،خیابان دانشگاه
-تقاطع شهدای ژاندارمری-پلاک ۱۵۸ساختمان دانشگاه-
طبقه دوم-واحد۴ تلفن ها: ۶۶۹۶۵۷۴۹-۲۲۰۸۵۱۱۱

فروشگاههای اینترنتی دیباگران تهران :
WWW.MFTBOOK.IR
www.dibagarantehran.com

سرشناسه:شفیعی رضوانی نژاد،نیما،۱۳۷۱-
عنوان و نام پدیدآور:وب کاوی با scrapy,beautifulsoup,selenium / مولفان:نیما شفیعی رضوانی نژاد،افشین اسمعیل زادآهندانی؛ ویراستار: مهدیه مخبری .
مشخصات نشر: تهران : دیباگران تهران :۱۴۰۱
مشخصات ظاهری:۱۵۲ ص:مصور،جدول
شابک: ۹۷۸-۶۲۲-۲۱۸-۶۴۹-۴
وضعیت فهرست نویسی: فیپا
موضوع:داده کاوی data mining
موضوع:اسکرپی scrapy(software framework)
موضوع:داده های کلان big data
شناسه افزوده: اسمعیل زادآهندانی،افشین،۱۳۷۱-
رده بندی کنگره:۷۶/۹ QA
رده بندی دیویی:۰۰۶/۳۱۲
شماره کتابشناسی ملی:۹۱۰۶۸۰۸

نشانی تلگرام:@mftbook نشانی اینستاگرام دیبا dibagaran_publishing

فهرست مطالب

مقدمه ناشر

خط مشی انتشارات مؤسسه فرهنگی هنری دیباگران تهران در عرصه کتاب هایی با کیفیت عالی است که بتواند خواسته های به روز جامعه فرهنگی و علمی کشور را تا حد امکان پوشش دهد.

هر کتاب دیباگران تهران، یک فرصت جدید شغلی و علمی

حمد و سپاس ایزد منان را که با الطاف بیکران خود این توفیق را به ما ارزانی داشت تا بتوانیم در راه ارتقای دانش عمومی و فرهنگی این مرز و بوم در زمینه چاپ و نشر کتب علمی و آموزشی گام‌هایی هرچند کوچک برداشته و در انجام رسالتی که بر عهده داریم، مؤثر واقع شویم.

گستردگی علوم وسرعت توسعه روزافزون آن، شرایطی را به وجود آورده که هر روز شاهد تحولات اساسی چشمگیری در سطح جهان هستیم. این گسترش و توسعه، نیاز به منابع مختلف از جمله کتاب را به عنوان قدیمی‌ترین و راحت‌ترین راه دستیابی به اطلاعات و اطلاع‌رسانی، بیش از پیش برجسته نموده است.

در این راستا، واحد انتشارات مؤسسه فرهنگی هنری دیباگران تهران با همکاری اساتید، مؤلفان، مترجمان، متخصصان، پژوهشگران و محققان در زمینه های گوناگون و مورد نیاز جامعه تلاش نموده برای رفع کمبودها و نیازهای موجود، منابعی پُربار، معتبر و با کیفیت مناسب در اختیار علاقمندان قرار دهد.

کتابی که در دست دارید تألیف **"جناب آقایان: نیما شفیعی رضوانی نژاد- افشین اسمعیل زاد آهندانی "** است که با تلاش همکاران ما در نشر دیباگران تهران منتشرگشته و شایسته است از یکایک این گرامیان تشکر و قدردانی کنیم.

با نظرات خود مشوق و راهنمای ما باشید

با ارائه نظرات و پیشنهادات وخواسته های خود،به ما کمک کنید تا بهتر و دقیق تر در جهت رفع نیازهای علمی و آموزشی کشورمان قدم برداریم.برای رساندن پیام هایتان به ما از رسانه های دیباگران تهران شامل سایتهای فروشگاهی و صفحه اینستاگرام و شماره های تماس که در صفحه شناسنامه کتاب آمده استفاده نمایید.

مدیر انتشارات

مؤسسه فرهنگی هنری دیباگران تهران
dibagaran@mftplus.com

مقدمه مولف

در سال های اخیر حجم داده های متنی که به وسیله وب در دسترس عموم قرار گرفته به طـور باورنکردنی افزایش یافته و تحلیل، بررسی و واکاوی این داده‌ها تا حد زیادی از عهده انسـان خارج شده و نیاز به یک نرم افزاری که بسیاری از کارهای تکراری و طاقت فرسای جمع‌آوری و تحلیل داده ها از وب را انجام بدهد بسیار احساس می‌شود. توانایی نوشتن یک ربات سـاده که داده ها را جمع‌آوری می‌کند و آن‌ها را در یک ترمینال پخش می‌کند یا آن‌هـا را در یـک پـا یگاه داده ذخیره می‌کند، اگرچه دشوار نیست، امـا همیشـه و حـس خاصـی را بـرای توسـعه دهندگان وب ایجاد می‌کند، مهم نیست که قبلا چند بار این کار را انجام داده باشید، چـون در هر پروژه کاری شما، شرایط متفاوت است و نیاز به تحلیل و بررسی دقیق‌تـر از اهـداف خـود دارید.

وقتی با برنامه نویسان در مورد وب اسکرپی صحبت می‌کنم، سوءتفاهم و سردرگمی زیادی در مورد این مورد وجود دارد. برخی از افراد مطمئن نیستند که قانونی است و یا نحوه برخـورد بـا مشکالتی همچون صفحاتی که با جاوا اسکریپت بارگزاری می‌شوند یا نوع رفتار اسـکرپی بـا صفحات ورود به سایت‌ها چگونه است. بسیاری در مورد چگونگی شروع یک پروژه بزرگ وب اسکرپی یا حتی محل پیدا کردن داده‌هایی که به دنبال آن هستند سردرگم هستند. این کتـاب به دنبال پایان دادن به بسیاری از این سؤالات رایج و تصـورات غلـط در مـورد وب اسـکرپی است که راهنمای جامعی برای اکثر پروژه‌های متداول وب اسکرپی ارائه می‌دهـد. ایـن حـوزه یک زمینه متنوع و سریع در حال تغییر است، و ما سعی کرده‌ایم مفاهیم سطح بالا و مثال‌هـای عینی را برای پوشش تقریبا هر پروژه جمع‌آوری داده ارائه دهیم.

در سراسر کتاب، نمونه کدهایی برای مفهوم هرچه دقیق تر مطالب ارائه شده است و به شـما امکان می‌دهد آن‌ها را پیاده‌سازی کنید.

نیما شفیعی رضوانی نژاد

افشین اسمعیل زاد