به نام خدا

انبار داده‌ها و داده‌کاوی

مؤلف

دکتر مهدی اسماعیلی

انبار داده ها و داده کاوی

مؤلف: مهدی اسماعیلی
ناشر: مؤسسه فرهنگی هنری دیباگران تهران
حروفچینی و صفحه‌آرایی: دیباگران تهران
طرح روی جلد: مجتبی حجازی
چاپ: درج عقیق
نوبت چاپ: چهارم
تاریخ نشر: ۱۳۹۸
تیراژ: ۵۰جلد
قیمت: ۱۰۰۰۰۰۰ریال
شابک: ۱-۴۳۹-۱۲۴-۶۰۰-۹۷۸
نشانی واحد فروش: تهران، میدان انقلاب، خ کارگر جنوبی، روبروی پاساژ مهستان، پلاک ۱۲۵۱
تلفن: ۶۶۴۱۰۰۴۶-۲۲۰۸۵۱۱۱
کد پستی: ۱۳۱۴۹۸۳۱۸۵
فروشگاههای اینترنتی:
www.mftbook.ir
www.dibagarantehran.com

نشانی تلگرام: @mftbook
نشانی اینستاگرام : Dibagaran_publishing
پست الکترونیکی: bookmarket@mftmail.com

سرشناسه : اسماعیلی ،مهدی،۱۳۵۰-
عنوان و نام پدید آور : انبار داده ها و داده کاوی/ مولف : مهدی اسماعیلی.
مشخصات نشر: تهران- دیباگران تهران -۱۳۹۶
مشخصات ظاهری:۴۱۶ ص. مصور.
شابک :۱-۴۳۹-۱۲۴-۶۰۰-۹۷۸
وضعیت فهرستنویسی : فیپا
یادداشت: کتابنامه :ص (۳۹۳)-۴۱۶
موضوع: داده کاوی
موضوع : data mining
موضوع: داده پردازی
موضوع : electronic data processing
موضوع : پایگاههای اطلاعاتی – مدیریت
موضوع : database management
رده بندی کنگره: ۱۳۹۵ ۵ الف ۲ د/۷۶۹ QA
رده بندی دیویی : ۰۰۶/۳۱۲
شماره کتابشناسی ملی : ۴۲۱۷۷۷۶

فهرست مطالب

مقدمه ناشر

خط مشی کیفیت انتشارات مؤسسه فرهنگی هنری دیباگران تهران در عرصه کتاب‌هایی است که بتواند خواسته‌های به روز جامعه فرهنگی و علمی کشور را تا حد امکان پوشش دهد.

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 دانشگاهی، علوم پایه و به ویژه علوم کامپیوتر و انفورماتیک گام‌هایی هرچند کوچک برداشته و در انجام رسالتی که بر عهده داریم، مؤثر واقع شویم.

گستردگی علوم و 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 روشن می‌نماید.

در این راستا، واحد انتشارات مؤسسه فرهنگی هنری دیباگران تهران با همکاری جمعی از اساتید، مؤلفان، مترجمان، متخصصان، پژوهشگران، محققان و نیز پرسنل ورزیده و ماهر در زمینه امور نشر درصدد هستند تا با تلاش‌های مستمر خود برای رفع کمبودها و نیازهای موجود، منابعی پُربار، معتبر و با کیفیت مناسب در اختیار علاقمندان قرار دهند.

کتابی که در دست دارید با همت آقای مهدی اسماعیلی و تلاش جمعی از همکاران انتشارات میسر گشته که شایسته است از یکایک این گرامیان تشکر و قدردانی کنیم.

کارشناسی و نظارت بر محتوا: زهره قزلباش

در خاتمه ضمن سپاسگزاری از شما دانش‌پژوه گرامی درخواست می‌نماید با مراجعه به آدرس dibagaran.mft.info (ارتباط با مشتری) فرم نظرسنجی را برای کتابی که در دست دارید تکمیل و ارسال نموده، انتشارات دیباگران تهران را که جلب رضایت و وفاداری مشتریان را هدف خود می‌داند، یاری فرمایید.

امیدواریم همواره بهتر از گذشته خدمات و محصولات خود را تقدیم حضورتان نماییم.

مدیر انتشارات

مؤسسه فرهنگی هنری دیباگران تهران
Publishing@mftmail.com

مقدمهٔ مولف

نه تو مانی، نه اندوه و نه هیچ یک از مردم این آبادی به تن لحظه خود جامه اندوه مپوشان هرگز

امروزه رشد چشمگیر داده‌ها در سازمان، مدیران را مجبور ساخته تا از ابزارهای مناسبی جهت تصمیم‌گیری‌های خود استفاده کنند. در چنین محیطی جهت تحلیل داده‌ها و در نهایت اتخاذ یک تصمیم مدیریتی مناسب، لازم است اطلاعات کلیه بخش‌ها جمع‌آوری و یکپارچه شوند. در این وضعیت انبار داده‌ها یک راه‌حل مناسب تلقی می‌شود. هدف اصلی یک انبار داده‌ها تسهیل در فرآیند تصمیم‌گیری و افزایش دانش افراد درگیر در این فرآیند است. اگرچه وجود انبار داده‌ها پیش‌نیاز تحلیل داده‌ها و همچنین داده‌کاوی نیست، ولی با داشتن یک انبار داده‌ها، عمل داده‌کاوی نیز بسیار آسان‌تر و مطمئن‌تر انجام خواهد شد. داده‌کاوی همچون هر کاوش و تحلیل دیگری، به دنبال گنجی از اطلاعات و دانش سودمند در میان اقیانوسی از مجموعه داده‌ها است و رشد بی‌رویه داده‌ها، اهمیت آن را دو چندان کرده است.

موضوع کتاب حاضر انبار داده‌ها و داده‌کاوی است و مطالب آن در ۱۰ فصل گردآوری و تنظیم شده است. فصل اول کتاب به معرفی مفاهیم پایه داده‌کاوی می‌پردازد. روش‌های آماده‌سازی داده‌ها موضوع فصل دوم کتاب است. بسیاری از صاحب‌نظران اهمیت پیش‌پردازش داده‌ها را بخش مهمی از فرآیند داده‌کاوی می‌دانند. فصل سوم کتاب به موضوع انبارش داده‌ها (ایجاد انبار داده‌ها) تخصیص داده شده است. در فصل‌های چهارم و پنجم، در مورد قوانین انجمنی بحث می‌شود. روش‌های دسته‌بندی موضوع فصل‌های ششم و هفتم کتاب را تشکیل می‌دهد. خوشه‌بندی و مباحث پیشرفته آن نیز به ترتیب موضوع‌های فصل‌های هشتم و نهم می‌باشد. فصل دهم کتاب به صورت خلاصه به موضوعاتی از قبیل متن‌کاوی، وِب‌کاوی، گراف‌کاوی و کاوش در میان داده‌های چندرسانه‌ای و فضایی می‌پردازد. بخش پایانی هر فصل نیز به معرفی منابع مربوط به موضوعات همان فصل می‌پردازد.

مطالب این کتاب به گونه‌ای نوشته شده‌اند تا خوانندگان محترم بتوانند مفاهیم آن را به راحتی درک کنند. واژه‌های معادل در این کتاب پیشنهادی هستند و چه بسا برای برخی از اصطلاحات، برابرهای دیگر (و احیاناً بهتر) بتوان یافت. آنچه مسلم است این است که اگرچه گاهی تبدیل واژه‌ها آنچنان که باید و شاید انجام نشده است، اما در ادای جملات و بیان موضوع تلاش فراوان شده است تا خوانندگان گرامی با متنی مبهم و گیج‌کننده روبرو نشوند.

در اینجا لازم می‌دانم از همه اساتید و دانشجویان به خاطر راهنمایی‌های ارزشمندشان در حین نگارش این کتاب سپاسگزاری کنم. همچنین از مدیریت محترم انتشارات دیباگران نیز به خاطر آماده‌سازی، چاپ و پخش این کتاب تشکر می‌کنم. رهین محبت بی‌دریغ خانواده‌ام هستم که با فراهم‌سازی محیطی مناسب مرا یار نمودند. با وجود همه سعی و تلاشی که در تمام مراحل آماده‌سازی این کتاب انجام گرفته است، یقین دارم که عاری از اشتباه نیست، چرا که تنها مکتوب بی‌نقص همان معجزه جاوید قرآن کریم است. در آخر ضمن سپاسگزاری از همه کسانی که مرا یاری داده‌اند و با پذیرش مسئولیت هرگونه کاستی احتمالی، امیدوارم که این اندک مفید افتد.

مهدی اسماعیلی
Msxpi2@yahoo.com