

رویکرد متن کاوی

نویسنده:

میر حسین سام دلیری



۱۳۹۸

سرشناسه: سام دلیری، امیرحسین، ۱۳۷۴-
عنوان و نام پدیدآور: رویکرد متن‌کاوی/نویسنده امیرحسین
سام‌دلیری.

مشخصات نشر: تهران: نوآوران دانش، ۱۳۹۹.

مشخصات ظاهری: ۸۵ص.: جدول، نمودار.

شابک: ۹۷۸-۶۲۲-۶۸۳۰-۶۸-۳

وضعیت فهرست‌نویسی: فیا

بازداشت: کتابنامه: ص. ۸۳-۸۵.

موضوع: داده‌کاوی

موضوع: Learning

رده‌بندی کنگر: ۲۸۰/۹

رده‌بندی دیویی: ۳۱۱/۰۰

شماره کتابشناسی ملی: ۷۱۵۰۹

عنوان: رویکرد متن‌کاوی

مؤلف: امیرحسین سام‌دلیری

طراح جلد: الویرا صیامی

نوبت چاپ: اول، ۱۳۹۹

شمارگان: ۱۰۰۰ جلد

قیمت: ۲۵۰۰۰ تومان

انتشارات نوآوران دانش

کلیه حقوق چاپ و نشر مخصوص و محفوظ ناشر است.

میدان انقلاب، ابتدای خیابان آزادی، تقاطع خیابان جمالزاده جنوبی، پلاک ۸۴، واحد ۴

تلفن مرکز پخش: ۶۶۵۶۸۹۱۷ - ۶۶۵۶۸۹۰۳ - ۶۶۵۶۴۶۰۸ - ۰۹۳۷۹۶۶۴۶۰۸

سفارش اینترنتی: www.novinelc.com



۷	پیشگفتار
۹	فصل اول: کلیات علوم داده و داده کاوی
۱۰	علوم داده
۱۲	داده کاوی
۱۵	کاوش متن (متن کاوی)
۱۶	فرآیند متن کاوی
۱۷	رسمهای کاوش متن
۱۸	پیش بینی آینده
۱۹	استخراج اطلاعات در زمان ماینر
۲۰	جنبه جدید بودن رویه کاوی
۲۳	فصل دوم: داده کاوی و مدل فوری اطلاعات
۲۴	مقدمه
۲۵	داده کاوی به عنوان سیر تکاملی فن آوری اطلاعات
۲۸	داده کاوی چیست؟
۳۱	چه نوعی از داده ها کاوش می شوند؟
۳۱	استخراج دانش
۳۵	انواع داده ها جهت داده کاوی
۳۷	داده کاوی و تکنیک های آن
۴۱	فصل سوم: بررسی کاوش متن
۴۲	قوانین انجمنی
۴۳	چالش های داده کاوی
۴۵	کاوش متن (کاوش متن)
۴۷	تفاوت کاوش متن و داده کاوی

۵۳.....	شبکه واژگان WordNet
۵۳.....	شبکه واژگان احساسی Senti-WordNet
۵۴.....	فرآیند کاوش متن
۵۵.....	فصل چهارم: استخراج دانش
۵۶.....	نمایش ویژگی‌ها
۵۹.....	استخراج دانش با استفاده از روش‌های یادگیری ماشین
۵۹.....	دقت طبقه‌بندی
۶۰.....	الگوریتم طاعنه‌بندی، نزدیکترین همسایه
۶۰.....	الگوریتم طبقه‌بندی، خط تصمیم
۶۱.....	طبقه‌بند نیویوز
۶۱.....	پژوهش‌های پیشین در حوزه کار با متن
۶۲.....	ارائه الگوریتمی جدید برای یادگیری انواع استخراج اطلاعات از متن
۶۵.....	فصل پنجم: مفاهیم اصلی مدل و نحوه اجرای فرآیند کاوش متن
۶۶.....	پیش پردازش داده‌ها
۶۸.....	معرفی نرم افزار Voyant
۷۲.....	ابزارهای استفاده شده در تحلیل
۷۷.....	طرح پراکش و خوشه بندی
۷۹.....	روند
۸۰.....	تحلیل روند و نتیجه گیری
۸۱.....	کلام پایانی
۸۳.....	منابع

پیشگفتار

پیش‌بینی علم و فناوری در پی‌ریزی برای نگاه به جلو و همچنین توجه به نیازهای آتی با تکیه بر نقش حیاتی علوم و فناوری در یک برنامه بلندمدت از اهمیت ویژه‌ای برخوردار است. پیش‌بینی فناوری نقطه شروع برنامه‌ریزی فناوری است که به‌عنوان یک ابزار مدیریتی می‌باشد. یکی از راهکارهای پیش‌بینی آینده، کاوش متون با استفاده از تکنیک‌های داده‌کاوی است. اگرچه پردازش خودکار اسناد با استفاده از تکنیک‌های مختلف اندازه‌گیری دهد تا عمق تحلیلی که انسان به آن اعتقاد دارد انجام شود، ولی این مبحث می‌تواند برای جداسازی نقاط کلیدی، طبقه‌بندی اسناد و تولید خلاصه‌های لازم به کار رود. خصوصاً در مواقعی که تحلیل دستی حجم زیادی از اسناد غیرممکن باشد. تحقیق بازار، فناوری هوش-کسب‌وکار، تحلیل درخواست‌ها و پرداخت‌های الکترونیکی تنها بخش از کاربردهایی هستند که کاوش متن می‌تواند در آن‌ها به کار گرفته شود.

فناوری اطلاعات یکی از عواملی است که در همبستگی جنبه‌های زندگی اجتماعی و انفرادی انسان تاثیر دارد و در سال‌های اخیر مورد توجه ویژه دولت‌ها قرار گرفته است. دولت‌ها برای جبران عقب ماندگی‌ها، ایجاد زندگی بهتر و نیز برای کسب قدرت بیشتر در رقابت‌های بین‌المللی و حفظ مزیت‌های خود به توسعه فناوری در حوزه‌های مختلف اقدام نموده‌اند. امروزه تحولات فناورانه تمام ابعاد سیاسی، اجتماعی، اقتصادی و فرهنگی را در نوردیده و به تناسب آن الگوهای رفتاری و روش‌های ارتباطی نیز دچار تحول شده‌اند.

در سازمانهای پیش‌نار، دانش اصلی‌ترین عامل در کسب مزیت رقابتی شناخته شده است. دنیا امروز به سمت به کارگیری سیستم‌های اثربخش‌تر حرکت می‌کند که بدون شک، دانش یکی از اجزای لاینفک چنین سیستم‌هایی است. سلسله مراتب مفاهیم تعریف‌کننده‌ی دانش به صورت زیر است:

- **داده‌ها:** کمیت‌های عددی یا غیرعددی هستند که از مشاهده، شهود یا محاسبه به دست می‌آیند.
- **اطلاعات:** مجموعه‌ای از داده‌ها همراه با جزئیات تشریحی، استنتاج‌ها و دیگر مواد متنی و نوشتاری در خصوص موضوع مطروحه یا فرآیندی خاص هستند.
- **فراداده‌ها:** شامل خلاصه‌های مشروح و تقسیم‌بندی سطح بالای داده‌ها و اطلاعات هستند.
- **دانش:** اطلاعات سازماندهی شده، ساختار یافته و خالص شده برای گسترش درک، آگاهی یا فهم است.
- **ادراک:** بر اثبات شفاف و کامل از ماهیت، اهمیت یا توضیح راجع به یک موضوع است.

این کتاب مسیری را در جهت ایجاد یک روش‌شناسی ساختارمند برای پیش‌بینی حوزه علم و فناوری با استفاده از رویکردها و تکنیک‌های جدید مبتنی بر خبرها و اسناد متونی ایجاد خواهد کرد که این بسترهای تاکنون در پژوهش‌های صورت گرفته مبتنی بر متون فارسی انجام نگرفته است.