

کلان داده و روش‌های جستجوی وب اجتماعی

تألیف:

پروفسور مایک ثلوال

ترجمه:

دکتر محمد حسن زاده
(رئیس دانشگاه تربیت مدرس)

محمد فلاح
(کارشناس ارشد، علم‌سنجی)

زهرا چمن
(کارشناس ارشد علم‌سنجی)



انتشارات شرکت آینده‌آموزان آتا

تهران

۱۳۹۷

سرشناسه: تلوال، مایکل آریجان Thelwall, Michael Arijan

عنوان و نام پدیدآور: کلان داده و روش‌های جستجوی وب اجتماعی / تألیف مایک تلوال؛ مترجمان محمدحسن زاده، محمد فلاح، زهرا چمنی.

مشخصات نشر: تهران: شرکت آینده آموزان آتا، ۱۳۹۷.

مشخصات ظاهری: ۳۳۶ ص.

شابک: ۹۷۸-۶۲۲-۹۹۹۶۹-۰۰۴

وضعیت فهرست‌نویسی: فیپا

یادداشت: عنوان اصلی: Big data and social webresearch methods

موضوع: شبکه‌های اجتماعی پیوسته

موضوع: Online social networks

موضوع: وب‌سنجی

موضوع: Webometri

موضوع: داده‌های کلان

موضوع: Big data

شناسه افزوده: حسن زاده، محمد، ۱۳۰۰ - مترجم

شناسه افزوده: فلاح، محمد، ۱۳۷۰ - مترجم

شناسه افزوده: چمنی، زهرا، ۱۳۸۰ - مترجم

رده‌بندی کنگره: ۱۳۹۷ ک ۸ ۴۲/۸ HM

رده‌بندی دیویی: ۳۰۲/۲۳۱

شماره کتابشناسی ملی: ۵۴۴۳۴۲۸

انتشارات شرکت آینده آموزان آتا

Big data and social webresearch methods

کلان داده و روش‌های جستجوی وب اجتماعی

مترجمان: محمدحسن زاده، محمد فلاح، زهرا چمنی

صفحه‌آرایی و طرح جلد: رضا حسن زاده و شاهو احمدیان

چاپ اول: زمستان ۱۳۹۷

شمارگان: ۵۰۰

قیمت: ۳۸۰۰۰ تومان

نشانی انتشارات: تهران، خیابان کارگر شمالی، خیابان گرد آفرید، نبش خیابان هیت، پلاک ۱۵، طبقه ۴- تلفن: ۰۲۱-۶۶۵۸۲۳۷۱

تلفن همراه: ۰۹۹۰۹۰۷۷۱۰۱ - پست الکترونیکی: info@artahub.com - آدرس سایت: www.artahub.com

کلیه حقوق این اثر برای ناشر محفوظ است و چنانچه هر شخص حقیقی یا حقوقی این اثر را بدون اجازه ناشر، چاپ یا تکثیر کند، مورد پیگرد قانونی قرار خواهد گرفت.

فهرست مطالب

عنوان	صفحه
پیشگفتار مترجم.....	۱۳
فصل اول: مقدمه و نظریه کلان داده.....	۱۵
پژوهش اطلاعات - محور: یک نظریه کلان داده.....	۱۸
بررسی اجزای.....	۱۹
فصل دوم: مطالعه شبکه ها و پیوندها.....	۲۰
مقدمه.....	۲۱
واژگان.....	۲۴
گردآوری اطلاعات شبکه در مورد افراد در وبگاه های اجتماعی.....	۲۵
جمع آوری اطلاعات شبکه.....	۳۲
ساختن شبکه از مجموعه ی وبگاه ها.....	۳۳
مصورسازی شبکه ها.....	۴۵
نمودارهای غیر شبکه ای برای نمایش شبکه ها.....	۴۹
اندازه گیری شبکه ها با تحلیلگر شبکه اجتماعی (SNA).....	۵۵
معیارهایی برای گره ها.....	۵۶
معیارهایی برای کل شبکه ها.....	۶۰
خلاصه و منابع پیشنهادی.....	۶۳
فصل سوم: تحلیل پیوند.....	۶۵
انواع تحلیل پیوند در وب سنجی.....	۶۸
سنجش های تأثیر پیوند.....	۷۰
تفسیر یافته ها.....	۷۴

۷۵.....	روش‌های جایگزین شمارش پیوند
۷۷.....	مطالعه موردی: پیوندهای مجله زیگزاگ
۷۹.....	تحلیل محتوای پیوندها
۸۱.....	ترسیم نقشه روابط پیوند
۸۵.....	مطالعات موردی
۹۰.....	ترسیم نقشه روابط هم‌پیوندی
۹۳.....	گزارش‌های تأثیر پیوند
۹۴.....	تحلیل پیوند میاس بزرگ با گروه‌های سایت چندگانه
۹۵.....	تفاوت‌های پیوند بین بخش‌ها: کاربرد در علم اطلاع‌رسانی
۹۸.....	خلاصه
۹۹.....	فصل چهارم: سنجش تأثیر وب
۱۰۳.....	سنجش تأثیر وب از طریق وب‌سنجی
۱۱۰.....	شاخص‌های سفارشی استاندارد وب
۱۱۲.....	تحلیل محتوا
۱۱۵.....	انتخاب مقوله (موضوع یا دسته‌های کلی)
۱۱۶.....	روش‌های نمونه‌گیری
۱۱۸.....	نمونه
۱۲۰.....	روایی و اعتبار
۱۲۱.....	تحلیل یو آر آل مربوط به توزیع نتایج
۱۲۵.....	تحلیل استنادی وب- کاربرد در علوم اطلاع‌رسانی
۱۲۸.....	مطالعات پیشرفته تأثیر وب
۱۲۹.....	خلاصه

فصل پنجم: توصیف انبوه شبکه‌ی وب	۱۳۰
مقدمه	۱۳۰
توصیفات انبوه شبکه‌ی وب: بنیاد و پایه	۱۳۱
نظری اجمالی بر روش‌ها	۱۳۳
WNTD خالص	۱۳۴
WNTD ترکیبی	۱۳۸
تحلیل محوایی ارتباطات: مشابه WNTD خالص	۱۴۰
مثال‌هایی از مطالعات WNTD	۱۴۲
مثال شماره یک: شبکه‌ی تویتر	۱۴۴
مثال ۲: پیوند وب یا شبکه‌های آمار	۱۵۳
خلاصه	۱۵۹
فصل ششم: تحلیل کامنت‌ها (نظرها) و عقاید	۱۶۱
مقدمه	۱۶۲
گردآوری کامنت‌ها جهت تحلیل	۱۶۵
واحدهای تحلیل	۱۶۵
استراتژی‌های نمونه‌گیری ابتدایی	۱۶۶
موضوعات نمونه‌گیری از طریق کلیدواژه‌ها	۱۶۹
ذخیره	۱۷۱
نمودارها و آمار ساده	۱۷۳
تحلیل محتوا	۱۷۸
برنامه‌ریزی یک تحلیل محتوا	۱۸۱
انجام تحلیل محتوا	۱۸۴

۱۸۷	گزارش نتایج
۱۸۹	خلاصه و پیشنهاد برای مطالعه
۱۸۹	نظریه داده بنیاد
۱۹۳	خلاصه و مواد خواندنی پیشنهادی
۱۹۴	تحلیل احساس
۲۰۰	سخن پایانی
۲۰۲	فصل هفتم: جستجوی وبلاگ و تویتر
۲۰۴	موتورهای جستجوی میکرو بلاگ
۲۰۶	جستجوهای تاریخ مشخص
۲۰۸	شناسایی الگو
۲۱۳	چک کردن نتایج تشخیص الگو
۲۱۵	محدودیت‌های داده‌های میکرو بلاگ
۲۱۷	فن‌های تحلیل بلاگ پیشرفته
۲۲۰	فن‌های تحلیل میکرو بلاگ پیشرفته
۲۲۲	خلاصه
۲۲۳	فصل هشتم: توصیف انبوه متن‌های وب
۲۲۳	مقدمه
۲۲۴	مرحله‌ی ۱: ایجاد و آزمایش جستارها
۲۲۷	مرحله‌ی ۲: اسپم‌های بعدی، مطالب غیر مرتبط و فیلترینگ دو نسخه‌ای
۲۳۲	مرحله‌ی ۳-۱: تحلیل محتوا
۲۳۳	مرحله‌ی ۳-۲: تحلیل‌های فراوانی نسبی واژگان در مقایسه با متون عمومی

مرحله‌ی ۳-۳: تحلیل‌های فراوانی نسبی واژگان برای مجموعه‌هایی از متون درون یک	
موضوع.....	۲۳۶
مرحله‌ی ۴-۳: سری‌های زمانی برای کلیدواژگان برای مجموعه‌های بزرگ متون.....	۲۳۸
مرحله‌ی ۵-۳: کاوش سری‌های زمانی برای مجموعه‌های بزرگ از متون.....	۲۴۰
مرحله‌ی ۶-۳: خلاصه و توصیف انبوه: یکپارچه کردن تحلیل‌ها.....	۲۴۱
مثال ۱: توت‌های بریتانیا درباره‌ی انقلاب مصر.....	۲۴۲
مثال ۲: توتسم.....	۲۴۹
خلاصه.....	۲۵۳
فصل نهم: جستجوهای رتور جستجوی خودکار: تحلیلگر وب‌سنجی.....	۲۵۴
مقدمه‌ای بر تحلیلگر وب‌سنجی.....	۲۵۶
گزارش‌های تأثیر وب تحلیلگر وب‌سنجی.....	۲۵۷
گزارش‌های تأثیر وب: مثال رابط کلاسیک.....	۲۶۱
نگاهی بر نتایج جستجو.....	۲۶۴
گزارش تأثیر پیوند تحلیلگر وب‌سنجی.....	۲۶۶
گزارش‌های تأثیر پیوند - مثال رابط کلاسیک.....	۲۶۸
نمودارهای تحلیلگر وب‌سنجی برای شبکه.....	۲۶۹
سازمان‌دهی مجدد، ذخیره‌سازی و چاپ نمودارهای شبکه.....	۲۷۰
نمودار شبکه - مثال رابط کلاسیک.....	۲۷۲
نمودارهای شبکه هم پیوند.....	۲۷۴
ویژگی‌های دیگر تحلیلگر وب‌سنجی.....	۲۷۵
فصل دهم: خزنده وب: سوسای بوت.....	۲۷۷
خزنده‌های وب.....	۲۷۸

۲۸۱	 مروری بر سوسای بوت
۲۸۲	 نمودارهای شبکه تعدادی وبگاه با سوسای بوت
۲۸۸	 کاربردهای دیگر خزنده وب
۲۹۰	 فصل یازدهم: موتورهای جستجو و پایایی داده‌ها
۲۹۰	 معمای موتورهای جستجو
۲۹۴	 حذف نتایج تکراری و نزدیک به تکراری
۲۹۶	 مقایسه موتورهای جستجوی مختلف
۲۹۶	 پژوهش رباره نتایج موتورهای جستجو
۲۹۹	 مدل‌سازی ساختار پیوندی وبی
۳۰۵	 فصل دوازدهم: رهگیری، درون‌خاک فعالیت‌های کاربران
۳۰۶	 تحلیل‌های وبی و تحلیل پوشش‌نگار کاربر
۳۰۸	 ارزیابی چندگانه وبگاه
۳۰۹	 تحلیل پرونده لاگ موتورهای جستجو
۳۱۱	 فصل سیزدهم: فنون پیشرفته
۳۱۱	 جداسازی عبارت جستجو
۳۱۹	 تحلیل موضوعات وبی
۳۲۱	 داده کاوی شبکه‌های اجتماعی
۳۲۳	 تحلیل شبکه‌های اجتماعی و جهان‌های کوچک
۳۲۵	 برچسب‌گذاری فولکسونومی
۳۲۶	 برنامه‌نویسی رابط‌های برنامه‌نویسی نرم‌افزارها (API) و مشاپ‌ها
۳۲۷	 خلاصه و رهنمودهایی برای آینده
۳۳۱	 واژه‌نامه

پیشگفتار مترجم

هم‌زمان با رشد روزافزون تولید اطلاعات و انتشار آن از طریق محمول‌های مختلف به‌ویژه سامانه‌ای برخط، داده‌های انبوه به‌صورت مستمر در تولید و منتشر می‌شود. افزایش حجم داده‌ای تولیدشده، اصطلاح کلان داده را به وجود آورده است. کلان داده به داده‌ای گفته می‌شود که حجم آن‌ها بسیار زیاد است و امکان تجزیه و تحلیل آن‌ها بدون به‌کارگیری سامانه‌های پیشرفته ممکن نیست. یکی دیگر از ویژگی‌های داده‌های کلان رتبه‌بندی چندگانه بین آن‌هاست که امکان تحلیل‌های چندبعدی را فراهم می‌آورد. انواع داده‌های کلان نیز متنوع و چندگانه است لذا تحلیل کلان داده با دشواری‌های مربوط به عدم تجانس نیز روبروست. از سوی دیگر، تحلیل کلان داده امکان تصمیم‌گیری دقیق‌تر را فراهم می‌آورد و بسترهای دستیابی به افق روشن در رابطه با پدیده‌ها و عملکرد متقاضیان آن‌ها را تسهیل می‌کند. با تحلیل دقیق کلان داده و کاربریست حساب‌شده آن‌ها، مدیریت امور بهبود می‌یابد، آینده‌نگری ارتقاء پیدا می‌کند و درنهایت عملکرد جمعی با بهره‌برداری از داده‌ها بهبود می‌شود.

از آنجایی که بیشترین داده‌های انبوه در بستر وب و شبکه‌های رایانه‌ای تولید می‌شود، یافتن راه‌هایی برای جستجو و تحلیل آن‌ها از اهمیت بسیار بالایی برخوردار است. پژوهشگران حوزه‌های مختلف علوم اجتماعی به‌ویژه علم اطلاعات و دانش‌شناسی به‌صورت روزمره با داده‌های انبوه مرتبط با موضوع خود سروکار دارند. تلاش برای

ارائه روش‌هایی متقن و حساب‌شده برای مواجهه با این داده‌ها از اهمیت بسیاری برخوردار است. مایک ثلوال یکی از نخستین دانشمندانی است که در حوزه تحلیل وب و اخیراً کلان داده به‌صورت عملیاتی وارد شده و نرم‌افزارهایی هم در این زمینه توسعه داده است.

کتاب «کلان داده و روش‌های جستجوی وب معنایی» در تداوم کتاب وب‌سنجی توسط مایک ثلوال به رشته تألیف درآمده است. از آنجایی که کتاب قبلی توسط مترجم، ترجمه و منتشر شده بود، به‌محض انتشار این کتاب، نیز اقدام برای ترجمه آغاز شد و طی نامه‌ای همانند کتاب قبلی اجازه ترجمه از مؤلف اخذ و کار ترجمه شروع شد. این کتاب در سیزده فصل ساماندهی شده و برای کلیه افرادی که در حوزه کلان داده فعالیت می‌کند، بسیار مفید خواهد بود. پژوهشگران حوزه علم اطلاعات و دانش‌شناسی، دانشجویان علوم اجتماعی، رسانه، علوم مدیریت، فناوری اطلاعات، جامعه‌شناسی و حوزه‌های مربوطه می‌توانند بهره‌های فراوانی از این کتاب ببرند. اکنون که این کتاب توسط انتشارات دانش‌بنیان آینده آموزان آتا مستقر در پارک علم و فناوری دانشگاه تربیت مدرس به رشته طبع آراسته شده است، بر خود لازم می‌دانم از مسئولان مربوطه و همکاران خود (جناب محمد فلاح و خانم زهرا چمنی) که زحمت بسیار زیادی در راه ترجمه و انتشار این اثر متحمل شده‌اند، قدردانی کنم و از خوانندگان ارجمند درخواست دارم که در صورت وجود هرگونه پیشنهاد اصلاحی مترجمان را از ایده‌ها و پیشنهادها خود آگاه سازند.

با آرزوی توفیق الهی

دکتر محمد حسن زاده

دانشیار دانشگاه تربیت مدرس