

وب کاوی

کاربردها و فنون

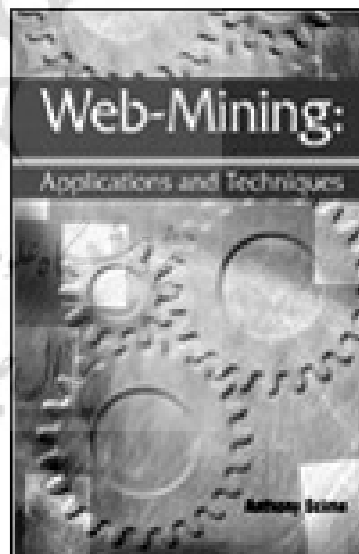
● حمید کشاورز

کارشناس ارشد کتابداری و اطلاع‌رسانی
کتابخانه ملی جمهوری اسلامی ایران

این رسانه اساساً ارتباطی است. مواردی از این قبیل که محیط وب به لحاظ محتوایی، ساختاری و کاربری از چه الگوهای پیروی می‌کند؛ چگونه می‌توان محتوای منابع اطلاعاتی موجود در وب را تحلیل کرد، چگونه که علاوه بر تأثیرگذاری بر الگوریتم‌های پیش‌بینی شده برای بازبازی منابع و نیز ساختاردهی به آنها، بتواند به کاربر در دستیابی به منابع اطلاعاتی مرتبط، دسته‌بندی شده و کارآمد کمک کند؛ چگونه می‌توان ساختار ارتباطی بین منابع اطلاعاتی مختلف را شناسایی و حتی ترسیم کرد؛ و اینکه چگونه می‌توان نوع کاربرد وب را از سوی کاربران شناسایی کرده و در خدمات و سیستم‌های آتی گنجانده؟

هر نوع پاسخی که به این پرسش‌ها داده شود، بر منافع دارندگان سایت‌ها و نیز کاربران آنها و در نهایت بر کیفیت ارائه خدمات در وب تأثیر خواهد گذاشت. برای یافتن پاسخ این پرسش‌ها باید اطلاعات موجود در محیط وب را بررسی و بازبینی کرد و بر اساس آن درک درستی از ویژگی‌های محتوایی، ساختاری و کاربری این محیط به دست آورد؛ چیزی که در ادبیات پژوهشی به «وب کاوی»^۱ مشهور شده است. این موضوع البته خود به بحث گسترده‌تری به نام داده کاوی^۲ که خود بخشی از فرآیند کشف دانش است، ارتباط دارد (اریناکی^۳، ۲۰۰۴، ۹، ۲). موضوع داده کاوی حدود چندین دهه قدمت دارد، یعنی از زمانی که سیستم‌های اطلاعاتی به وجود آمده‌اند، وجود دارند. اما با ظهور وب این موضوع گسترش بیشتری یافت و به حوزه‌های کوچک‌تری از جمله وب کاوی تقسیم شد. از سوی دیگر، وب، مفاهیم سنتی بازبازی اطلاعات و فنون مربوط به آن را به چالش کشانده است (شنکر^۴ و دیگران، ۲۰۰۵، ص ۸). به عبارت دیگر، خصوصیات مدارک موجود در وب، سبک و میزان ارائه مدارک در وب و از همه مهم‌تر پویایی وب، الگوریتم‌های بازبازی اطلاعات را که بر تطابق کلیدواژه‌ای و دسته‌بندی در فضای برداری^۵ مبتنی بود، ناکارآمد ساخته است.

در نتیجه عواملی نظیر آنچه در پیش آمد، متخصصان روش‌های



■ Scime, Anthony. (Ed.) (2005). **Web Mining: Applications and Techniques**. Idea Group Inc. 442 pp. ISBN: 1-59140-414-2

یکی از مهم‌ترین رسانه‌های اطلاعاتی و ارتباطی که در حال حاضر به فراوانی مورد استفاده قرار می‌گیرد، شبکه جهانی وب است. محیط وب ویژگی‌های گوناگونی دارد که از جمله برجسته‌ترین آنها باید به گستردگی، گوناگونی منابع و پویایی آن اشاره کرد. با رشد سریع وب هر یک از این ویژگی‌ها جنبه‌های گوناگون و بیشتری نیز پیدا می‌کنند. اما آنچه اهمیت فوق‌العاده‌ای دارد، دریافتن قوانین حاکم بر



**با توجه به گستردگی منابع و خدمات
وب، وب‌کاوی خود به سه حوزه
تقریباً جداگانه تقسیم شده است:
محتواکاوی وب، ساختار کاوی وب و
کاربرد کاوی وب**

نویسی را در پیش گرفته‌اند تا بتوانند کمبودهای موجود در بازیابی اطلاعات را مرتفع ساخته و الگوهای ساختاری، محتوایی و کاربری موجود در وب را به‌خوبی نشان دهند. در نتیجه چنین بحث‌هایی وب‌کاوی به‌عنوان موضوعی مستقل و قابل پژوهش در دهه اخیر مطرح شده است. اگرچه متخصصانی از حوزه رایانه پیش‌تاز چنین بحث‌هایی بوده‌اند، اما به‌دلیل گستردگی استفاده از وب در امور مختلف تجاری، ارتباطی و اطلاع‌رسانی و مانند آن وب‌کاوی به بحث متخصصان سایر حوزه‌ها نیز تبدیل شده است.

وب‌کاوی که اولین بار از سوی اتریبونی^۱ (۱۹۹۷) به‌کار رفته، چنین تعریف شده است: «کاربرد فنون داده‌کاوی برای کشف و استخراج خودکار اطلاعات از منابع و خدمات وب» (کسالا و بلوکل^۲، ۲۰۰۰، ص ۲؛ شنکر و دیگران، ۲۰۰۵، ۳؛ برنت، هوتو و استم^۳، ۲۰۰۲، ۲۲۶؛ فاکا و لانزی^۴، ۲۰۰۳، ۱۴۰؛ کولی، مباشر و سربواستاو^۱، ۱۹۹۷، ۱).

با توجه به گستردگی منابع و خدمات وب، وب‌کاوی خود به سه حوزه تقریباً جداگانه تقسیم شده است: محتواکاوی وب^۱، ساختار کاوی وب^۲، و کاربرد کاوی وب^۳ (کسالا و بلوکل، ۲۰۰۰، ص ۲؛ برنت، هوتو و استم، ۲۰۰۲، ص ۲۶۵؛ اربناکی، ۲۰۰۴، ۲). در هر یک از این حوزه‌ها موضوعات متفاوتی مورد توجه قرار می‌گیرد. هدف محتواکاوی، کشف اطلاعات مفید از مدارک و مطالب موجود در وب است. اگرچه مطالب موجود در مدارک وب بسیار متنوع هستند مانند متن، تصویر، صدا، ابرداده و پیوندهای فرامتنی^۴، در محتواکاوی ساختار منابع و نحوه کاربرد آنها کمتر مورد توجه قرار می‌گیرد. در ساختار کاوی ساختار موجود بین مدارک و زنجیره ارتباطی آنها از طریق تحلیل پیوندهای فرامتنی مطالعه می‌شود. در کاربرد کاوی نوع کاربرد مدارک و

به‌طور کلی نحوه استفاده از وب از سوی کاربران بررسی شده و الگوی استفاده مشخص می‌گردد. این امر با مطالعه گزارش‌های^{۱۵} تولیدشده از سوی کاربران که در انواع سرویس‌دهنده‌ها ذخیره می‌گردد، صورت می‌گیرد. درحالی‌که در محتواکاوی و ساختار کاوی اطلاعات دسته اول مورد استفاده قرار می‌گیرد، در کاربرد کاوی اطلاعات دسته دوم یعنی گزارش‌های موجود در سرویس‌دهنده‌ها مورد استفاده پژوهشگران قرار می‌گیرد. کاربرد کاوی کمی دیرتر از دو روش دیگر مورد توجه قرار گرفته است (شنکر و دیگران، ۲۰۰۵، ۴). هرچه وب در مسیر تبدیل شدن به رسانه‌ای تجاری، ارتباطی و آموزشی به‌پیش می‌رود، مطالعه دقیق نوع کاربرد آن دارای اهمیت بیشتری می‌شود. البته لازم به یادآوری است که نمی‌توان مرز روشی بین این سه دسته از پژوهش‌ها قرار داد و در پژوهش‌های مختلف می‌توان با ترکیب آنها به نتایج معتبرتری دست یافت.

کتاب وب‌کاوی: کاربردها و فنون با هدف نشان‌دادن این

هرچه وب در مسیر تبدیل شدن به رسانه‌های تجاری، ارتباطی و آموزشی به پیش می‌رود، مطالعه دقیق نوع کاربرد آن دارای اهمیت بیشتری می‌شود

مطالب کتاب به پنج بخش تقسیم شده است. بخش دوم کتاب، «محتواکاوی وب»، فصل‌های سوم تا نهم را شامل می‌شود. در فصل سوم، «کاربرد اطلاعات محیطی برای ساخت یک سیستم کاوش مصنوعی»، چندین قاعده مهم برای ارزیابی ارتباط صفحات وب با یک موضوع خاص توصیف می‌شوند که مهم‌ترین آنها نظم رتبه‌ای بر اساس ترسیم ربط محیطی^{۱۸} است. بدین مفهوم که چگونه می‌توان ارتباط بین صفحات مختلف ناظر بر موضوعی خاص را در قالب نموداری ترسیم و رتبه‌بندی کرد.

در فصل چهارم، «یادگیری هستی‌شناسی از یک ساختار وب حوزه‌ای»، یک روش شناسی برای شناخت هستی‌شناسی‌ها از یک ساختار ویژه وب معرفی می‌گردد. کاربرد این روش شناسی که به استخراج واژگان^{۱۹} منجر می‌شود، نوعی تفسیر معنایی^{۲۰} از واژه‌های مرتبط را فراهم می‌کند و به‌طور خودکار هستی‌شناسی حوزه‌ای را در پی خواهد داشت.

در فصل پنجم، مارس^{۲۱} معرفی می‌شود. مقاله از یک الگوریتم تطبیق افزایشی پرده برمی‌دارد که به کاربر در بازیابی مدارک موردعلاقه‌اش کمک می‌کند. این الگوریتم از یک استراتژی گسترش پرسش^{۲۲} کاربر برای بهبود و اجرای مجدد پرسش برای شناخت علاقه کاربر استفاده می‌کند. به‌عبارت دیگر، این الگوریتم با بررسی پرسش‌های کاربر از سیستم، واژه‌های مرتبط با آنها را به‌طور خودکار شناسایی کرده و با گسترش فزاینده دامنه پرسش درصد پاسخ‌گویی به نیاز کاربر برمی‌آید.

فصل ششم، «کاربرد عناوین برای داده‌کاوی وب»، یک رویکرد غیرمستقیم برای نمایه‌سازی منابع چندرسانه‌ای موجود در صفحات وب با استفاده از عنوان آنها را پیشنهاد می‌کند. این رویکرد از این‌رو حائز اهمیت است که درک متن نسبت به چندرسانه‌ای‌ها برای موتورهای جست‌وجو بسیار ساده‌تر است و عناوین صفحات معمولاً به محتوای مدارک اشاره دارند.

در فصل هفتم، «به سوی نظریه بحران برخاسته از سیستم امنیتی مصنوعی برای وب‌کاوی»، بخشی از یک پروژه بزرگ‌تر معرفی می‌شود و آن ساخت یک سیستم محتواکاوی پویای وب است. در این مقاله سیستم ایمنی مصنوعی برای طبقه‌بندی نامه‌های الکترونیکی معرفی می‌گردد که با استفاده از آن نامه‌های الکترونیکی با توجه به محتوایشان طبقه‌بندی شده و برای تأمین امنیت آنها تدابیر خاصی اندیشیده می‌شود.

در فصل هشتم، «معانی ایکس.ام. ال»^{۲۳}، نیز یک روش جدید برای افزودن معنا به مدارک ایکس.ام. ال با اضافه کردن اطلاعات

حوزه پژوهشی نوظهور و اهمیت آن به چاپ رسیده است. هرچند تأکید اصلی آن کاربرد فنون وب‌کاوی در بهبود جست‌وجو عنوان شده است (پیشگفتار کتاب) و همین امر باعث جذابیت بیشتر این کتاب و نیز پذیرش آن از سوی خوانندگانی غیر از علوم رایانه‌ای و سیستم‌های اطلاعاتی مانند ارتباطات، تجارت الکترونیکی و کتابداری و اطلاع‌رسانی خواهد شد.

کتاب شامل ۱۹ مقاله از نویسندگان برجسته حوزه وب‌کاوی است که آنتونی اسکیم^{۱۶} آنها را گردآوری آورده است. از نکات جالب این کتاب تعداد نویسندگان مقالات آن است که به ۴۰ نفر می‌رسند. این امر حاکی از ماهیت اینگونه پژوهش‌هاست که غالباً از درون پروژه‌های مختلف مایه می‌گیرند. نکته قابل توجه دیگر حضور نویسندگانی با ملیت‌های آسیای شرقی و نیز مسلمان است که گواه علاقه‌مندی آنان به انجام پژوهش‌هایی از این دست می‌باشد.

نویسنده مقاله اول، «مدیریت ابرداده: الزامی برای انباشت و مدیریت دانش»^{۱۷}، به اهمیت کاربرد ابرداده‌ها در مدیریت محتوای مدارک موجود در وب می‌پردازد. از این‌رو که سازمان‌ها و افراد مختلف برای معرفی خدمات خود از واژه‌های مختلفی استفاده می‌کنند، در این مقاله استفاده از ابرداده‌های استاندارد برای معرفی خدمات توصیه می‌شود. در طول مقاله به اهمیت وب‌کاوی در تأمین خواست کاربران و نیز دارندگان سایت‌ها اشاره شده است.

یکی از مباحث مهمی که در چند سال اخیر در پژوهش‌های وب‌کاوی مورد توجه ویژه قرار گرفته، شخصی‌سازی است. شخصی‌سازی در پی آن است تا با شناخت ویژگی‌ها و نیازهای کاربران، خدمات خاص آنها را فراهم کند. در حقیقت با استفاده از فنون شخصی‌سازی، افراد مختلف از خدمات ویژه خودشان بهره‌مند خواهند شد. در فصل دوم، «کاوش برای شخصی‌سازی وب»^{۱۸} نویسندگان مقدمه‌ای درباره این بحث مهم ارائه می‌کنند. البته در طول کتاب در بخش‌های ۳، ۵، ۱۱، ۱۳، ۱۴، ۱۷، ۱۸ و ۱۹ نیز بحث شخصی‌سازی مطرح است که گویای اهمیتی است که این موضوع در پژوهش‌های وب‌کاوی دارد. دو فصل اول و دوم با هدف تهیه مقدمه‌ای برای بخش‌های بعدی کتاب به‌صورت بخش مستقلی در ابتدای کتاب آورده شده‌اند، اما به‌دلیل محتوای تقریباً فنی آنها به‌نظر می‌رسد که خوانندگان مبتدی با خواندن آنها نتوانند به‌آسانی به مطالعه بخش‌های بعدی کتاب بپردازند.

به‌تبعیبت از دسته‌بندی رایج، در ادامه فصل‌های مختلف کتاب در سه بخش جداگانه محتواکاوی وب، ساختارکاوی وب و کاربردکاوی وب گنجانده شده‌اند که با احتساب بخش نتیجه‌گیری و مقدمه،



کتاب در ارتباط مستقیم با حوزه‌های پژوهشی و کاربردی است که با ارائه خدمات مبتنی بر وب همچون سیستم‌های اطلاعاتی، بازیابی اطلاعات و تجارت الکترونیک سروکار دارند

با در نظر گرفتن وب به‌مثابه محیطی که در آن مدارک مختلف با هم ارتباط دارند، می‌توان آن را به یک نمودار متشکل از زنجیره‌های ارتباطی بین منابع تشبیه کرد و بر اساس آن منحنی‌های مختلفی ترسیم کرد. در این فصل به خوشه‌بندی و دسته‌بندی این نمودارها برای ساده‌سازی فرآیندهای نمایش و جست‌وجو اشاره شده است.

در بخش سوم که فصل‌های سیزدهم تا هجدهم را شامل می‌شود، به کاربرد کاوی وب اختصاص دارد. در فصل سیزدهم، «یکپارچه‌سازی معنایی و کاربرد کاوی وب برای شخصی‌سازی»، از ملزومات مورد نیاز برای یکپارچه‌سازی دانش معنایی موجود در منابع مختلف از جمله محتوا و ساختار وب‌سایت‌ها بحث شده است. آنچه در این فصل ارائه شده است، یک چارچوب عمومی برای یکپارچه‌سازی کامل هستی‌شناسی‌های مصنوعی و فنون کاربرد کاوی و شخصی‌سازی وب با استفاده از پیش‌پردازی^{۲۸} و کشف الگو می‌باشد.

در فصل چهاردهم بیتز^{۲۹} از صاحب‌نظران حوزه بازیابی اطلاعات و ویراستار کتاب ارزشمند بازیابی نوین اطلاعات به اهمیت بررسی گزارش‌های تولیدشده و ذخیره‌شده در سرویس‌دهنده‌ها و نیز

معناشناختی به نشانه‌های عناصر این زبان مورد بحث قرار می‌گیرد. در فصل نهم، «طبقه‌بندی بر اساس مکعب داده‌ای»^{۳۰}، روش جدیدی برای سازماندهی فایل‌های موجود در پایگاه‌های اطلاعاتی معرفی می‌شود که در آن سه طبقه‌بند^{۳۱} جدید روی یک مکعب داده‌ای برای اهداف مبادلاتی و تحلیلی طراحی می‌شوند.

بخش دوم کتاب، «ساختار کاوی وب»، شامل فصل‌های دهم تا دوازدهم است. در فصل دهم، «پالایش و اعتباربخشی داده‌ها برای تحلیل چندجانبه ساختاری پیوند سایت‌ها»، چندین روش پالایش و اعتباربخشی به داده‌های مربوط به پیوندها برای استفاده در ساختار کاوی وب توصیف شده‌اند. از آنجا که نویسنده مقاله، مایک تلسوال^{۳۲} از پژوهشگران فعال در حوزه وب‌سنجی و ساختار کاوی وب بوده و خود کتاب تحلیل پیوند: رویکرد اطلاع‌رسانی را تألیف کرده، توانسته است فنون کاربردی جدیدی را در زمینه پژوهش‌های مرتبط با تحلیل ساختار موجود بین مدارک وب معرفی کند. وی در این مقاله از روش تحلیل چندجانبه ساختار پیوند سایت‌ها برای شناخت الگوهای موجود در مجموعه‌های مشخصی از سایت‌های وب سخن گفته است.

در فصل یازدهم، «استخراج و سفارش اطلاعات با استفاده از کارگزاران متعدد»، از چالش‌های پیش‌رو در محیط‌های پیچیده و مسائل مرتبط با سرریز اطلاعات^{۳۳} سخن به میان آمده است. برای مواجهه با این چالش‌ها، نویسنده الگوی استفاده از کارگزاران متعدد در سیستم‌های بازیابی را پیشنهاد می‌کند.

در فصل بعدی، «خوشه‌بندی نموداری وب برای نمایش و گردش در فضای مجازی»، به موضوع ترسیم نموداری ساختار وب و چگونگی خوشه‌بندی این نمودارها پرداخته شده است. به عبارت دیگر،



با در نظر گرفتن وب به مثابه محیطی که در آن مدارک مختلف با هم ارتباط دارند، می توان آن را به یک نمودار متشکل از زنجیره های ارتباطی بین منابع تشبیه کرد

در آخرین فصل از بخش دوم کتاب، «کاربرد کاوی وب: الگوریتم ها و نتایج» بار دیگر بر باز کاوی گزارش های دسترسی وب، تحلیل الگوریتم های پردازش و استخراج دانش از چنین گزارش هایی تأکید شده و نیز فنونی برای باز کاوی گزارش ها به صورتی جامع پیشنهاد می شود.

بخش پنجم کتاب نتیجه گیری نام دارد. اما بر خلاف تصور در آن مقاله ای همچون سایر مقالات گنجانده شده و ارزیابی کلی از مطالب کتاب صورت نگرفته است. در این مقاله با عنوان «رایحه یک گروه خبری: فراهم آوری دسترسی مشخص به سایت های یوزنت از طریق وب کاوی» کاربرد فنون وب کاوی در مسیر برطرف کردن مسائل مرتبط با دسترسی شخصی به خدمات یوزنت مورد ملاحظه قرار گرفته است. در این مقاله سه حوزه اصلی وب کاوی تحلیل شده است: محتوا کاوی که در آن تأکید ویژه ای بر کشف موضوع می شود، ساختار کاوی که در آن ساختار خبرهای یوزنت با کاربرد فنون ساختار کاوی مورد بررسی قرار می گیرد و کاربرد کاوی که در آن فنونی برای ردگیری^{۳۳} و پرونده سازی^{۳۴} کاربران بر اساس مطالعه گزارش های وب به کار گرفته می شود. این فصل با مروری بر فنون شخصی سازی و نیز توصیف شخصی سازی ویژه دسترسی به یوزنت به پایان می رسد.

از نقاط قوت کتاب باید عنوان آن را ذکر کرد. زیرا به خوبی از محتوای کتاب خبر می دهد. به علاوه پیروی از دسته بندی رایج شیوه های وب کاوی و گنجاندن فصل های مختلف کتاب زیر هریک از این دسته ها به خواننده در درک مفاهیم و نیز فنون کاربرد آنها در پژوهش های مختلف کمک شایانی می کند. دیگر اینکه کتاب

پرسش هایی که از سوی کاربران ارائه شده، پرداخته شده است. در طول مقاله با توجه فراوان به بحث کاربرد کاوی وب خاطر نشان شده است که باید پرسش ها و رفتار کاربر به دقت تحلیل شود تا بتوان از نتایج این تحلیل در طراحی و بهبود رابط های کاربری موتورهای جستجو، عملکرد نمایه سازی و رتبه بندی نتایج استفاده کرد.

فصل پانزدهم به یکی از موضوعات مهم در حوزه تحلیل گزارش های وب، حجم فراوان داده های ناظر بر رفتاری مشخص و تکراری از سوی جستجوی کاربر، پرداخته است. از آنجاکه کاربران در جستجوی مسیری تقریباً مشخص و مشابهی را دنبال می کنند، در گزارش های وب حجم انبوهی از داده ها به این رفتارها اختصاص دارد و این امر تحلیل گزارش ها را با مشکل مواجه می سازد. در این فصل، «وب کاوی کارآمد برای الگوهای مسیری پیمایی»^{۳۵} نویسندگان از دو الگوریتم جدید برای حل این مسئله سخن گفته اند.

فصل شانزدهم، «تحلیلی از الگوهای مشاهده مدارک کاربران موتورهای جستجوی وب» که از جانب دو تن از پژوهشگران حوزه اطلاع رسانی به نگارش درآمده، الگوی مشاهده صفحات نتایج جستجو و سایر صفحات وب توسط کاربران مورد توجه قرار می گیرد. در این فصل از ارزش کاربرد تحلیل های سنتی گزارش در شناسایی این الگوها بحث شده است.

در فصل هفدهم، «معماری نرم افزار مبتنی بر جاوا توزیعی برای کاربرد کاوی وب» تکنیکی از هریک از رویکردهای مختلف کاربرد کاوی وب مانند خوشه سازی، قوانین همراهی^{۳۶} (همراهی عبارات و واژه ها در مدارک مختلف) و الگوی بی درپی سخن گفته شده است. این الگوها در معماری یادگیرنده با نام متالا^{۳۷} یکپارچه می شوند.

23. XML Semantics
24. Classification on Top of Data Cube
25. Classifier
26. Mike Thelwall
27. Information overload
28. Preprocessing
29. Yates
30. Efficient Web Mining For Traversal Path Patterns
31. Association rules
32. METALA
33. Tracking
34. Profiling

منابع و مأخذ

1. Berendt, B., Hotho, A. and Stumme, G. (2002). **Towards Semantic Web Mining. International Semantic Web Conference**, pp. 264-278
2. Cooley, R. Mobasher, B. and Srivastava, J. (1997) "Web Mining: Information and Pattern Discovery on the World Wide Web" In Proceedings of the 9th IEEE International Conference on Tool with Artificial Intelligence. [Online] Available at: maya.cs.depaul.edu/~mobasher/papers/webminer-tai97.pdf (Accessed 23 October 2007).
3. Eirinaki, M. (2004). Web Mining: A Roadmap, Technical Report, DB-NET 2004, available at <http://www.engr.sjsu.edu/meirinaki/papers/NE-MIS.pdf> (accessed 23 October 2007).
4. Etzioni, O. (1996). "The World-Wide Web: quagmire or gold mine?", **Communications of the ACM**, v.39 n.11, p.65-68.
5. Facca, F. M. and Lanzi P. L. (2003). "Recent Developments in Web Usage Mining Research, In Kambayashi Y. et al., **Data Warehousing and Knowledge Discovery**": 5th International Conference, DaWak 2003 Prague, September 3-5, 2003 Proceedings, Czech Republic, page 140-150.
6. Kosala, R, and Blockeel, H. (2000). "Web mining research: A survey". **SIGKDD explorations**, 2(1):1-15.
7. Schenker, A., Bunke, H., Last, M. and Kandel, A. (2005). "Graph-Theoretic Techniques for Web Content Mining" World Scientific Publishing Company.

مجموعه‌ای از مقالات نویسندگان و پژوهشگران فعال حوزه وب کاوی است و خواننده می‌تواند با نگاهی به زندگینامه آنها در پایان کتاب از گذشته آنان و نیز کیفیت پروژه‌های در دست انجام آگاه شود و به‌طور کلی دیدگاه جامعی نسبت به موضوع کتاب پیدا کند.

از آنجاکه کتاب مجموعه‌ای از مقالات مختلف است، معایبی که برای اینگونه از کتاب‌ها متصور است برای کتاب حاضر نیز از قبیل نداشتن یکپارچگی در ارائه مطالب، گنجاندن فصل‌ها در چارچوبی از پیش تعریف‌شده یا اساساً چارچوبی ساختگی، در نظر نگرفتن نوع مخاطب و میزان دانش آنها، کاربرد فزاینده واژه‌های تخصصی و ماهیت عملگرایی مقالات وجود دارد.

با این همه کتاب از جمله نخستین تلاش‌ها برای معرفی این حوزه پژوهشی نوظهور، میان‌رشته‌ای و مهم به‌شمار می‌رود. کتاب در ارتباط مستقیم با حوزه‌های پژوهشی و کاربردی است که با ارائه خدمات مبتنی بر وب همچون سیستم‌های اطلاعاتی، بازیابی اطلاعات و تجارت الکترونیک سروکار دارند.

کتاب دارای یک نمایه پایانی است که فقط چهار صفحه است و برای بازنامهی مطالب کتابی با موضوعات جدید و ۴۴۲ صفحه خالی از اشکال نیست. به‌نظر می‌رسد در تهیه آن شتاب‌زده عمل شده است؛ امری که ظاهراً کل کتاب نیز با آن دست به‌گریبان است.

پی‌نوشت‌ها

1. Web mining
2. Data mining
3. Eirinaki
4. Schenker
5. Vector Space Model
6. Etzioni
7. Kosala and Blockeel
8. Berendt, Hotho and Stumme
9. Facca and Lanzi
10. Cooley, Mobasher and Srivastava
11. Web content mining
12. Web structure mining
13. Web usage mining
14. Hyperlinks
15. Logs
16. Anthony Scime
17. Metadata Management: A Requirement for Web Warehousing and Knowledge Management
18. Relevancy context graph
19. Terminology
20. Semantic interpretation
21. MARS: Multiplicative Adaptive Refinement Web Search
22. Query expansion