



## وب کاوی و کاربرد آن در صنعت

ولی الله سرلک<sup>۱</sup>، میترا گودرزی<sup>۲</sup>، مهدی رائیجی یانه سری<sup>۳</sup>، میر مرسل مدنی<sup>۴</sup>، سپیده سلطانی فهرج<sup>۵</sup>

<sup>۱</sup> دانشجوی کارشناسی ارشد مهندسی کامپیوتر، هوش مصنوعی، دانشگاه بین المللی امام رضا (ع) مشهد  
 Mohandesprg@gmail.com  
<sup>۲</sup> کارشناس مهندسی فناوری اطلاعات، دانشگاه پیام نور الیگودرز  
 Engmg.engprg@gmail.com

<sup>۳</sup> دانشجوی کارشناسی ارشد مهندسی کامپیوتر، هوش مصنوعی، دانشگاه بین المللی امام رضا (ع) مشهد  
 Mehdi.Rayeji@yahoo.com  
<sup>۴</sup> عضو هیئت علمی دانشگاه آزاد اسلامی گرگان واحد کردکوی  
 m.madani@kordkuyiau.ac.ir

<sup>۵</sup> دانشجوی کارشناسی ارشد مهندسی کامپیوتر، هوش مصنوعی، دانشگاه بین المللی امام رضا (ع) مشهد  
 Sepideh.soltanifahraj@yahoo.com

**چکیده :** با افزایش چشمگیر حجم اطلاعات و توسعه وب، نیاز بهروش ها و تکنیک هایی که بتوانند امکان دستیابی کارا به داده ها و استخراج اطلاعات از آنها را فراهم کنند، بیش از پیش احساس می شود. وب کاوی یکی از زمینه های تحقیقاتی است که با به کارگیری تکنیک های داده کاوی به کشف و استخراج خودکار اطلاعات از استناد و سرویس های وب می پردازد. در واقع وب کاوی، فرآیند کشف اطلاعات و دانش ناشناخته و مفید از داده های وب می باشد. روش های وب کاوی بر اساس آن که چه نوع داده ای را مورد کاوش قرار می دهد، به سه دسته کاوش محتوای وب، کاوش ساختار وب و کاوش استفاده از وب تقسیم می شوند. طی این گزارش پس از معرفی وب کاوی و بررسی مراحل آن، ارتباط وب کاوی با سایر زمینه های تحقیقاتی بررسی شده و به چالش ها، مشکلات و کاربردهای این زمینه تحقیقاتی اشاره می شود. همچنین هر یک از انواع وب کاوی به تفصیل مورد بررسی قرار می گیرند. برای این منظور مدل ها، الگوریتم ها و کاربردهای هر طبقه معرفی می شوند.

**کلمات کلیدی :** وب کاوی، کاربردهای وب کاوی، انواع وب کاوی، وب کاوی در تجارت الکترونیک، وب کاوی و زمینه های تحقیقاتی

### ۱ - مقدمه

با توسعه سیستم های اطلاعاتی، داده به یکی از منابع پراهمیت سازمان ها مبدل گشته است. بنابراین روش ها و تکنیک هایی برای دستیابی کارا به داده، اشتراک داده، استخراج اطلاعات از داده و استفاده از این اطلاعات، مورد نیاز می باشد. با ایجاد و گسترش وب و افزایش چشمگیر حجم اطلاعات، نیاز به این روش ها و تکنیک های بیش از پیش احساس می شود. وب، محیطی وسیع، متنوع و پویا است که کاربران متعدد اسناد خود را در آن منتشر می کنند. در حال حاضر بیش از دو بیلیون صفحه در وب موجود است و این تعداد با نرخ ۷.۳ میلیون صفحه در روز افزایش می یابد. با توجه به حجم وسیع اطلاعات در وب، مدیریت آن با ابزارهای سنتی تقریباً غیر ممکن است و ابزارها و روش هایی نو برای مدیریت آن مورد نیاز است.

### ۲ - مطلب اصلی

با توجه به گسترش روز افزون حجم اطلاعات در وب و ارتباط وب کاوی با تجارت الکترونیکی، وب کاوی به یک زمینه تحقیقاتی وسیع مبدل گشته است. طی این گزارش پس از بررسی مراحل وب کاوی، انواع آن معرفی می شوند. سپس ارتباط وب کاوی با سایر زمینه های تحقیقاتی بررسی شده و به چالش ها و مشکلات این زمینه تحقیقاتی اشاره می شود. در ادامه هر یک از انواع وب کاوی به تفصیل مورد بررسی قرار می گیرند. برای این منظور مدل ها، الگوریتم ها و کاربردهای هر طبقه معرفی می شوند. در پایان نیز به برخی از نمونه کاربردهای واقعی وب کاوی اشاره می شود.

وب کاوی شامل چهار مرحله اصلی می باشد:

۱. پیدا کردن منبع: این مرحله شامل بازیابی استناد و ب مورد نظر می باشد.
  ۲. انتخاب اطلاعات و پیش پردازش: در این مرحله به صورت خودکار اطلاعات خاصی از استناد بازیابی شده، انتخاب و پیش پردازش می شوند.
  ۳. تعمیم: در این مرحله الگوهای عام در یک یا چندین سایت وب کشف می شود.
  ۴. تحلیل: در این مرحله الگوهای به دست آمده در مرحله قبل اعتبار سنجی و تفسیر می شوند.
- در مرحله اول داده ها از منابع موجود در وب مانند خبرنامه های الکترونیکی، گروه های خبری، استناد HTML، پایگاه داده های متنی و ... بازیابی می شوند. مرحله انتخاب و پیش پردازش شامل هر گونه فرآیند تبدیل داده های بازیابی شده در مرحله قبل می باشد. این پیش پردازش می تواند کاهش کلمات به ریشه آنها، حذف کلمات زائد، پیدا کردن عبارات موجود در متن و تبدیل بازنمایی داده ها به قالب رابطه ای یا منطق مرتبه اول باشد. در مرحله سوم از تکنیک های داده کاوی و یادگیری ماشین برای تعمیم استفاده می شود. همچنین باید توجه داشت که کاربران نقش مهمی در فرآیند استخراج اطلاعات و دانش از وب ایفا می کنند. این نکته به ویژه در مرحله چهارم از اهمیت بسزایی برخوردار است. به این ترتیب وب کاوی، فرآیند کشف اطلاعات و دانش ناشناخته و مفید از داده های وب می باشد. این فرآیند به طور ضمنی شامل فرآیند کشف دانش در پایگاه داده ها (KDD<sup>۵</sup>) نیز می شود. در واقع وب کاوی گونه توسعه یافته KDD است که بر روی داده های وب عمل می کند.

### ۳ - وب کاوی و زمینه های تحقیقاتی مرتبط

وب کاوی با زمینه های مختلف تحقیقاتی علوم کامپیوتر همچون داده کاوی، پایگاه داده، بازیابی اطلاعات، هوش مصنوعی، یادگیری ماشین، پردازش زبان طبیعی، استخراج اطلاعات، انبار داده ها، طراحی واسطه کاربر و ... در ارتباط تنگاتنگ است. در این بخش ارتباط این زمینه تحقیقاتی با برخی از زمینه های مرتبط بررسی می شود.

#### ۱-۱ - وب کاوی و داده کاوی

وب کاوی و داده کاوی ارتباط بسیار نزدیکی با یکدیگر دارند. داده کاوی فرآیند ارائه پرس و جوها و استخراج الگوهای پرس و جوها و اطلاعات مفید و ناشناخته از داده هایی است که معمولاً در پایگاه داده ها ذخیره شده اند. در واقع بسیاری از تکنیک های داده کاوی قابل استفاده در وب کاوی هستند. اما حوزه وب کاوی وسیع تر از داده کاوی است و این دو زمینه تحقیقاتی در جنبه های مختلفی از یکدیگر متفاوتند که برخی از آنها عبارتند از: در داده کاوی، داده ها ساخت یافته هستند و معمولاً در پایگاه داده ها وجود دارند. اما در وب، داده ها عموماً غیر ساخت یافته هستند. جمع آوری و مدیریت داده ها در وب دشوار است.

داده ها در وب تنها شامل محتوای مستندات و صفحات وب نیستند. بلکه در وب دو نوع داده اصلی دیگر نیز برای کاوش مورد استفاده قرار می گیرند. نوع اول، اطلاعات ساختاری وب است که منظور از آن پیوندهای بین صفحات وب می باشد. نوع دوم نیز، اطلاعات مربوط به نحوه استفاده کاربران از وب است. در واقع تحلیل رفتار کاربر در استفاده از وب، ترجیحات و علائق وی درباره نوع و قالب اطلاعات، ... بخش مهمی از وب کاوی است. در داده کاوی این دو نوع از داده وجود ندارند. مسئله دیگری که در وب کاوی مطرح است، حفظ حریم کاربران<sup>۶</sup> است. تکنیک های داده کاوی معمولاً در یک محیط بسته به کار می روند. در حالی که تکنیک های وب کاوی در محیط باز وب انجام می شوند. بنابراین باید تضمین شود، اطلاعات شخصی و خصوصی کاربران مورد سوء استفاده قرار نمی گیرند

## ۲-۳ - وب کاوی و بازیابی اطلاعات

بعضی محققین معتقدند که کشف منبع یا سند (بازیابی اطلاعات) در وب، نمونه ای از وب کاوی است و برخی وب کاوی را مرتبط با بازیابی اطلاعات هوشمند می دانند. منظور از بازیابی اطلاعات، بازیابی خودکار اسناد مرتبط و در عین حال بازیابی کمترین حد ممکن از اسناد غیر مرتبط می باشد. اهداف اصلی بازیابی اطلاعات شاخص گذاری<sup>۱</sup> متون و جستجو برای اسناد مرتبط در یک مجموعه می باشد. در حال حاضر تحقیقات در زمینه بازیابی اطلاعات شامل مدلسازی، طبقه بندی<sup>۲</sup> اسناد، واسطه های کاربری، تصویر سازی داده<sup>۳</sup>، جداسازی<sup>۴</sup> و ... می باشد. آنچه در این میان می تواند به عنوان نمونه ای از وب کاوی در نظر گرفته شود، طبقه بندی اسناد است که در شاخص گذاری مورد استفاده قرار می گیرد. با چنین دیدگاهی وب کاوی به بخشی از فرآیند بازیابی اطلاعات مبدل می گردد

## ۳-۱ - وب کاوی و استخراج اطلاعات

هدف از استخراج اطلاعات تبدیل مجموعه ای از اسناد به اطلاعات خلاصه شده و تحلیل شده می باشد. در حالی که تمرکز اصلی در بازیابی اطلاعات بر انتخاب اسناد مرتبط است، استخراج اطلاعات بر استخراج واقعی مرتبط از اسناد تکیه دارد. همچنین در استخراج اطلاعات، ساختار یا بازنمایی یک سند مد نظر قرار می گیرد، در حالی که در بازیابی اطلاعات، یک سند مجموعه ای نامربت از کلمات است. ساخت یک سیستم استخراج اطلاعات برای محیط پویا و متنوعی چون وب امکان پذیر نیست و بیشتر سیستم های ایجاد شده بر سایت های وب خاصی متمرکز می شوند. برخی دیگر از سیستم های استخراج اطلاعات از تکنیک های یادگیری ماشین و داده کاوی برای یادگیری قوانین و الگوهای استخراج استفاده می کنند. با این دیدگاه، وب کاوی بخشی از فرآیند استخراج اطلاعات می باشد.

البته دیدگاه های دیگری درباره ارتباط این دو وجود دارد. برخی معتقدند استخراج اطلاعات گونه ای از مرحله پیش پردازش (مرحله بعد از بازیابی اطلاعات و قبل از اعمال تکنیک های داده کاوی) در فرآیند وب کاوی می باشد.

به طور کلی دو گونه متفاوت از استخراج اطلاعات وجود دارد. استخراج اطلاعات از متون غیر ساخت یافته و استخراج اطلاعات از داده های نیمه ساخت یافته، برای استخراج اطلاعات از متون غیر ساخت یافته معمولاً نوعی پیش پردازش زبانی قبل از به کارگیری تکنیک های داده کاوی استفاده می شود. بنابراین این نوع از استخراج اطلاعات ارتباط نزدیکی با تکنیک های پردازش زبان طبیعی دارد. اما با ایجاد و گسترش وب نیاز به روش های بازیابی اطلاعات از متون ساخت یافته می باشد. استخراج اطلاعات ساخت یافته متفاوت از استخراج اطلاعات غیر ساخت یافته است، چرا که معمولاً از اطلاعاتی مانند تگ های HTML، جدا کننده ها، ... استفاده می کند. بیشتر روش های ساخت یافته ای که در وب به کار می روند، از تکنیک های یادگیری ماشین برای استخراج قوانین استفاده می کنند.

## ۴-۱ - وب کاوی و یادگیری ماشین

وب کاوی معادل یادگیری از وب یا به کارگیری تکنیک های یادگیری ماشین در وب نیست. کاربردهایی از یادگیری ماشین در وب وجود دارد که نمونه های وب کاوی به شمار نمی آیند. یک مثال از این نوع کاربردها، به کارگیری تکنیک های یادگیری ماشین برای یافتن بهترین مسیر در پیمایش وب توسط Spider ها است.

از طرف دیگر علاوه بر تکنیک ها و روش های یادگیری ماشین، روش های دیگری هم برای وب کاوی به کار می رود. به عنوان مثال الگوریتم های اختصاصی برای یافتن Hub ها و Authority ها در وب وجود دارد. (مفهوم Hub و Authority در بخش ۷ معرفی خواهد شد.) با این حال ارتباط نزدیکی بین یادگیری ماشین و وب کاوی وجود دارد. در واقع تکنیک های یادگیری ماشین از وب کاوی پشتیبانی می کنند و قابل استفاده در فرآیندهای وب کاوی می باشند. به عنوان مثال تحقیقات نشان می دهد استفاده از تکنیک های یادگیری ماشین در طبقه بندی اسناد، می تواند دقیق طبقه بندی را در مقایسه با استفاده از روش های سنتی بازیابی اطلاعات افزایش دهد.

## ۴ - انواع وب کاوی

- روش های وب کاوی بر اساس آن که چه نوع داده ای را مورد کاوش قرار می دهند، به سه دسته تقسیم می شوند:
  - کاوش محتوای وب<sup>۵</sup>: کاوش محتوای وب فرآیند استخراج اطلاعات مفید از محتوای مستندات وب است. محتوای یک سند وب متناظر با مفاهیمی است که آن سند در صدد انتقال آن به کاربران است. این محتوا می تواند شامل متن، تصویر، ویدئو، صدا و یا رکوردهای ساخت یافته مانند لیست ها و جداول باشد. در این میان کاوش متن بیش از سایر زمینه ها مورد تحقیق قرار گرفته است. از جمله این تحقیقات می توان به تشخیص موضوع<sup>۶</sup>، استخراج الگوهای ارتباط<sup>۷</sup>، خوشه بندی<sup>۸</sup> و طبقه بندی اسناد و ب اشاره کرد. روش ها و تکنیک های موجود در این گروه، از تکنیک های بازیابی اطلاعات و پردازش زبان طبیعی نیز استفاده می

کنند. هر چند در پردازش تصویر و بینایی ماشین تحقیقات زیادی در زمینه استخراج دانش از تصاویر انجام شده است، اما به کارگیری این تکنیک‌ها در کاوش محتوای وب چندان چشمگیر نبوده است.

- کاوش ساختار وب<sup>۶</sup>: وب را می‌توان به صورت گرافی که گره‌های آن اسناد و یال‌های آن پیوندهای<sup>۷</sup> بین اسناد است، بازنمایی کرد. کاوش ساختار وب، فرآیند استخراج اطلاعات ساختاری از وب می‌باشد.
- کاوش استفاده از وب<sup>۸</sup>: کاوش استفاده از وب، کاربرد تکنیک‌های داده کاوی برای کشف الگوهای استفاده از وب، به منظور درک و برآوردن بهتر نیازهای کاربران می‌باشد. این نوع از وب کاوی، داده‌های مربوط به استفاده کاربران از وب را مورد کاوش قرار می‌دهد.

باید توجه داشت که مرز مشخصی میان سه گروه وب کاوی وجود ندارد. به عنوان مثال تکنیک‌های کاوش محتوای وب می‌توانند علاوه بر به کارگیری متن مستندات، از اطلاعات کاربران هم استفاده کنند. همچنین می‌توان از ترکیب تکنیک‌های فوق برای حاصل شدن نتایج بهتر استفاده کرد.

## ۵ کاربردهای وب کاوی

تکنیک‌ها و روش‌های وب کاوی از کاربرد وسیعی در حوزه‌های مختلف همچون تجارت الکترونیکی، دولت الکترونیکی، آموزش الکترونیکی، آموزش از راه دور، سازمان‌های مجازی، مدیریت دانش، کتابخانه‌های دیجیتال، ... برخوردارند. در این قسمت برخی از حوزه‌های کاربردی وب کاوی و نمونه کاربردهای موفقی که از تکنیک‌های وب کاوی استفاده کرده‌اند، معرفی می‌شوند.

### ۱-- تجارت الکترونیکی

تکنیک‌های وب کاوی می‌توانند تا حد چشمگیری به موفقیت تجارت الکترونیکی کمک نمایند. اگر سازمانی بخواهد از طریق وب فعالیت تجاری کند، باید تصمیم بگیرد چه نوع داده‌ای را در وب قرار دهد. وب کاوی می‌تواند در تشخیص نوع این داده‌ها به سازمان‌ها کمک کند. برای مثال سایتی که کتاب می‌فروشد، باید روش مناسبی برای تبلیغ کتب مختلف انتخاب کند. برای این منظور این سایت می‌تواند، بر اساس علایق کاربران، آن‌ها را دسته بندی کرده و تبلیغات مربوط به یک کتاب خاص را برای گروه کاربران مربوطه انجام دهد. به عنوان مثال دیگر، اگر سازمانی قصد تجارت با سازمان دیگری را داشته باشد، می‌تواند از وب کاوی برای یافتن شریک تجاری مناسب و تشخیص رقبا استفاده کند. همچنین می‌تواند خط مژی رقبا را تعیین کرده و بر اساس آن سرویس‌ها و کالاهای خود را قیمت گذاری کند.

قبل از آن که سازمانی از وب کاوی استفاده کند، باید زمینه‌هایی را که استفاده از وب کاوی در آن متمرث مر خواهد بود، تعیین کند. سپس به جمع آوری داده در آن زمینه خاص می‌پردازد. مثلاً داده‌های مربوط به کاربران را جمع آوری می‌کند. مرحله بعد، تعیین نوع کاوش است. به عنوان مثال می‌توان کاربران و رقبا را دسته بندی کرد و یا ارتباط بین کاربران، رقبا و شرکا را استخراج نمود. در پایان نیز باید روش‌های برای اعتبارسنجی نتایج به دست آمده معرفی شود.

یک نمونه از سایت‌های تجاری که از تکنیک‌های وب کاوی همچون استخراج ارتباط بین صفحات مشاهده شده توسط کاربر، تحلیل مسیر کلیک شده، ... استفاده می‌کند، سایت Amazon است که موفقیت‌های چشمگیری برای آن به دنبال داشته است.

### ۲- موتورهای جستجو

بیشتر موتورهای جستجو برای بازیابی صفحات مرتبط با پرس‌وجوی کاربر از محتوای صفحات وب استفاده می‌کنند. Google اولین موتور جستجویی بوده است که از ساختار پیوندها در وب و کاوش آن برای بازیابی و رتبه بندی صفحات استفاده کرده است. برای این منظور Google از الگوریتم Page Rank که در بخش‌های قبل معرفی شد، استفاده می‌کند. همچنین Google دارای این قابلیت است که اطلاعات مربوط به جریان کلیک‌های کاربر را نگهداری کند. این اطلاعات استفاده از وب، برای بهبود نتایج بازیابی شده و سفارشی سازی آن‌ها به کار می‌رود. از دیگر سرویس‌های Google که در آن از تکنیک‌های وب کاوی استفاده می‌شود، می‌توان به Google News اشاره کرد. این سرویس اخبار موجود در روزنامه‌های مختلف وب را جمع آوری کرده و سپس با استفاده از روش‌های طبقه بندی و خوش بندی آن‌ها را در گروه‌ها و دسته‌های مختلف سازماندهی می‌کند.

## ۵ - ۴ - حراجی در وب

سایت ebay یکی از موفق ترین سایت های تجاری وب است که امکان به حراج گذاشتن کالاهای مختلف را فراهم می کند. اما مatasfane این خاصیت وب که اشخاص در آن می توانند در آن اطلاعات واقعی خود را ارائه نکنند، برای ebay مشکلاتی به وجود آورده است، چرا که نمی توان بین پیشنهادهای واقعی و غیر واقعی تمایز قائل شد. برای حل این مشکل از تکنیک های وب کاوی استفاده کرده است. برای این منظور رفتار و الگوهای ارائه کردن پیشنهاد توسط شرکت کنندگان در حراجی، تحلیل می شود تا الگوی پیشنهادات غیر واقعی مشخص و با آن برخورد مناسب گردد.

## ۶ - نتیجه گیری

با ایجاد و گسترش وب و افزایش چشمگیر حجم اطلاعات، روش ها و تکنیک هایی برای استفاده از این اطلاعات و استخراج اطلاعات جدید از آنها مورد نیاز می باشد. روش های سنتی بازیابی اطلاعات که برای جستجوی اطلاعات در پایگاه داده ها به کار می روند، قابل استفاده در وب نمی باشند. وب کاوی که به کارگیری تکنیک های داده کاوی برای کشف و استخراج خودکار اطلاعات از اسناد و سرویس های وب می باشد، می تواند برای این منظور به کار رود.

وب کاوی شامل چهار مرحله اصلی پیدا کردن منبع، انتخاب اطلاعات و پیش پردازش، تعمیم، تحلیل می باشد. روش های وب کاوی بر اساس آن که چه نوع داده ای را مورد کاوش قرار می دهد، به سه دسته تقسیم می شوند: کاوش محتوای وب، کاوش ساختار وب کاوش استفاده از وب. تکنیک ها و روش های وب کاوی از کاربرد وسیعی در حوزه های مختلف همچون تجارت الکترونیکی، دولت الکترونیکی، آموزش الکترونیکی، آموزش از راه دور، سازمان های مجازی، مدیریت دانش، کتابخانه های دیجیتال، ... برخوردارند. البته وب کاوی با چالش ها و محدودیت های متنوعی روبرو است. و در حال حاضر تحقیقات بسیاری در زمینه وب کاوی در حال انجام است که هدف آن ها حل این مشکلات می باشد.

## مراجع

- [1] [1] R. Baeza-Yates. "Web mining in search engines". Proceedings of the 27th conference on Australasian computer science, Volume 26, 2004, 3-4.
- [2] [2] K. Bharat, B. Chang, M. Henzinger, M. Ruhl. "Who links to whom: Mining linkage between web sites". In IEEE International Conference on Data Mining (ICDM '01), San Jose, California, November 2001.
- [3] [3] S. Chakrabarti, B. Dom, D. Gibson, J. Kleinberg, SR Kumar, P. Raghavan, S. Rajagopalan, A. Tomkins. "Mining the link structure of the World Wide Web". IEEE Computer, 32(8), August 1999, 60-67.
- [4] [4] S. Chakrabarti, B. Dom, P. Indyk, "Enhanced hypertext categorization using hyperlinks", SIGMOD 1998, 307-318.
- [5] [5] S. Chakrabarti. "Data mining for hypertext: A tutorial survey". SIGKDD Explorations, 1(2),2000, 1-11.
- [6] [6] R. Cooley, B. Mobasher, and J. Srivastava. "Web Mining: Information and Pattern Discovery on the World Wide Web". In Proceedings of the 9th IEEE International Conference on Tools with Artificial Intelligence (ICTAI'97), November 1997.
- [7] [7] R. Cooley, "The use of Web structure and content to identify subjectively interesting Web usage patterns". ACM Trans. Internet Technology, 3(2), 2003, 93-116.
- [8] [8] V. Crescenzi, P. Merialdo, P. Missier. "Clustering Web pages based on their structure". Data Knowl. Eng. 54(3), 2005, 279-299.
- [9] [9] P. Desikan, J. Srivastava, V. Kumar, P.-N. Tan, "Hyperlink Analysis -- Techniques & Applications", Army High Performance Computing Center Technical Report, 2002.
- [10] [10] D. Donato, S. Leonardi, S. Millozzi, and P. Tsaparas. "Mining the inner structure of the web graph". In Proceedings of the Eighth International Workshop on the Web and Databases (WebDB), June 2005, 145-150.
- [11] [11] F. Facca, P. Lanzi. "Mining interesting knowledge from weblogs: a survey". Data Knowl. Eng. 53(3), 2005, 225-241.
- [12] [12] J. Furnkranz. Web mining. The Data Mining and Knowledge Discovery Handbook, pages. Springer, 2005, 899- 920
- [13] [13] S. Garofalakis, R. Rastogi, S. Seshadri, K. Shim. "Data mining and the web: past, present and future". In 2nd ACM International Workshop on Web Information and Data Management (WIDM), 1999, 43-47.
- [14] [14] E. Glover, K. Tsoutsouliklis, S. Lawrence, D. Pennock, G. Flake. "Using web structure for classifying and describing web pages". In International World Wide Web Conference, May 2002.
- [15] [15] M. Henzinger. "Hyperlink analysis for the web". IEEE Internet Computing, 5(1), January/February 2001, 45-50.
- [16] [16] R. Kosala, H. Blockeel. "Web mining research : A survey". In SIGKDD Explorations Newsletter of the ACM Special Interest Group on Knowledge Discovery and Data Mining, 2(1), 2000, 1-15.
- [17] [17] B. Liu, K. Chang, "Editorial: Special Issue on Web Content Mining". SIGKDD Explorations special issue on Web Content Mining, Dec, 2004.
- [18] [18] B. Mobasher, N. Jain, E.H. Han, and J. Srivastava. "Web mining: Pattern discovery from world wide web transactions". Technical Report TR-96050, Department of Computer Science, University of Minnesota, M inneapolis, 1996.
- [19] [19] L. Page, S. Brin, R. Motwani, T. Winograd. "The PageRank Citation Ranking: Bringing Order to the Web". Stanford Publications, 1998

- 
- <sup>1</sup> Generalization
  - <sup>2</sup> Validation
  - <sup>3</sup> Stemming
  - <sup>4</sup> Stop Words
  - <sup>5</sup> Knowledge Discovery in Data Base
  - <sup>6</sup> Data Warehouse
  - <sup>7</sup> Privacy
  - <sup>8</sup> Indexing
  - <sup>9</sup> Classification
  - <sup>10</sup> Data Visualization
  - <sup>11</sup> Filtering
  - <sup>12</sup> Web Content Mining
  - <sup>13</sup> Topic Discovery
  - <sup>14</sup> Association Pattern
  - <sup>15</sup> Clustering
  - <sup>16</sup> Web Structured Mining
  - <sup>17</sup> Hyperlink
  - <sup>18</sup> Web Usage Mining

This document was created with Win2PDF available at <http://www.daneprairie.com>.  
The unregistered version of Win2PDF is for evaluation or non-commercial use only.