

سورة الاحقاف

سمینار

عنوان

بررسی روش‌های خوشه‌بندی کاربران وب

چکیده.....	۵
------------	---

فصل اول: مقدمه

۱-۱- مقدمه	۷
------------------	---

فصل دوم: بستر تحقیق

۱-۲- مقدمه	۱۰
۲-۲- خوشه بندی	۱۱
۳-۲- هدف از خوشه بندی	۱۲
۴-۲- تفاوت خوشه بندی با طبقه بندی	۱۳
۵-۲- نظریه مجموعه های فازی	۱۴
۶-۲- مفهوم خوشه بندی فازی	۱۴
۱-۶-۲- تفاوت خوشه بندی فازی با خوشه بندی کلاسیک	۱۵
۷-۲- وب کاوی	۱۶
۱-۷-۲- چالش های وب کاوی	۱۷
۲-۷-۲- اجزای اصلی وب کاوی	۱۸
۸-۲- خوشه بندی کاربران وب	۱۸
۹-۲- جمع بندی	۱۹

فصل سوم: پیشینه تحقیق

۱-۳- مقدمه	۲۱
۲-۳- روش خوشه بندی فازی FCM	۲۱
۳-۳- الگوریتم های پایه ای خوشه بندی فازی	۲۳
۴-۳- الگوریتم خوشه بندی C میانگین	۲۳
۵-۳- روش های ارائه شده برای خوشه بندی کاربران وب	۲۷
۶-۳- معیارهای کارایی خوشه بندی کاربران وب	۲۸
۷-۳- جمع بندی	۳۴

فصل چهارم: نتیجه گیری و پیشنهادات

۴-۱- نتیجه گیری و پیشنهادات ۳۶

مراجع ۳۸

فهرست اشکال

- شکل ۱-۲ - خوشه‌بندی نمونه‌های ورودی ۱۱
- شکل ۲-۲ - خوشه بندی وسایل نقلیه ۱۲
- شکل ۳-۲ - مجموعه داده پروانه‌ای ۱۵
- شکل ۴-۲ - خوشه بندی فازی داده ۱۶
- شکل ۱-۳ - توزیع یک بعدی نمونه ها ۲۵
- شکل ۲-۳ - خوشه بندی کلاسیک نمونه های ورودی ۲۵
- شکل ۳-۳ - خوشه بندی فازی نمونه ها ۲۶

چکیده

خوشه‌بندی یکی از روش‌های رایج مورد استفاده در حوزه کاوش استفاده از وب به منظور خوشه‌بندی جریان کلیک کاربران است. به طور کلی کاربران وب رفتارهای متفاوتی متناسب با نیازهای اطلاعاتی و وظایف مورد علاقه خود در بازدید از وبسایت‌ها از خود نشان می‌دهند، تمامی رفتارهای پیمایشی کاربران در فایل‌های ثبت وب قابل ردیابی است. یکی از تکنیک‌های مورد استفاده در کاربرد کاوی وب خوشه‌بندی کاربران وب می‌باشد. در تکنیک خوشه‌بندی کاربرانی که رفتار پیمایشی مشابهی دارند در یک خوشه قرار می‌گیرند. هر خوشه منجر به ایجاد پروفایل‌های کاربری می‌شود که در برنامه‌های کاربردی مانند واکنشی اولیه و حافظه نهان مورد استفاده قرار می‌گیرد. تکنیک‌های متداول و استاندارد کاربرد کاوی وب برای خوشه‌بندی کاربران وب می‌تواند الگوهای کاربردی را مستقیماً کشف کند، اما این تکنیک‌ها به‌طور خودکار نمی‌توانند مشخصات یا کیفیت عامل‌های پنهانی که منجر به کشف الگوهای پیمایشی مشترک می‌شوند را تعیین کنند. بنابراین نیاز به گسترش تکنیک‌هایی می‌باشد تا بتوان بصورت خودکار اهداف اساسی پیمایشی کاربران را شناسایی و رابطه معنایی پنهان میان کاربران وب و همچنین رابطه معنایی پنهان بین کاربران وب و اشیاء وب را استخراج کرد.

کلمات کلیدی: کاربران وب، خوشه بندی، رفتار، فازی، وقایع وب

فصل اول

مقدمه

با وجود زیر ساخت‌ای تولید علم در جامعه نظیر سیستم‌های کامپیوتری و اینترنت امروزه دیگر تولید اطلاعات کار سهل الوصول و فراگیری می‌باشد طوریکه هر شخصی به راحتی می‌تواند آنچه می‌خواهد را با دیگران در فضای مجازی به اشتراک بگذارد و تمامی لحظات خود را ثبت نماید. تولید اطلاعات از این دست با سرعت بسیاری در حال رشد می‌باشد اما همه داده‌های موجود دارای ارزش اطلاعاتی نیستند. هنگامیکه کاربر تقاضای خود را در اینترنت دنبال می‌کند برای رسیدن به جواب مطلوب پس از صرف زمان و حجم داده که در کل آنرا هزینه می‌نامیم به جواب مورد نظرش برسد [1]. برای کاستن از این هزینه تحقیقات بسیاری صورت گرفته که نتیجه تمام آنها را می‌توان به طور خلاصه بدین صورت بیان کرد که اگر بتوانیم کاربران را در هنگام ورود به سایت یا پورتال یا فروشگاه مجازی یا صفحه آغازین موتور جستجوگر یا غیره به طریق مختلف شناسایی کنیم و کلیه فعالیت‌های آنها شامل حرکت چشم، لینک‌های کلیک شده، موارد جستجو شده، کلمات کلیدی مورد نظر و هر آنچه که به رفتار آنها در تعامل با پیشنهادات و نتایجی که از طرف ما با آن روبرو می‌شود را در یک پایگاه ثبت کنیم می‌توانیم با استفاده از الگوریتم‌های شناخته شده کاربران را به خوشه‌ها و دسته‌های مختلفی تقسیم کنیم بطوریکه اعضاء هر کلاس و گروه بیشترین شباهت را با یکدیگر و بیشترین اختلاف را با اعضاء گروه‌ها و کلاس‌های دیگر داشته باشند. اما انجام این عمل نیازمند این است که در ابتدا نتایج و موارد مورد تقاضا کاربران را بدون در نظر گرفتن سابقه قبلی آنها و صرفاً از روی وابستگی کلمات کلیدی به یکدیگر و با الگوریتم‌های خوشه‌بندی به دسته‌های مختلف تقسیم کنیم. سپس با شروع عکس‌العمل از طرف کاربران به نتایج داده شده باید اقدام به ثبت رفتارها و لینک‌های مورد علاقه کاربر و نقاط مورد نظر صفحه و غیره کنیم تا بتوانیم از روی این اطلاعات به سلیقه و هدف اصلی کاربر دست پیدا کنیم که البته برای ثبت این اطلاعات می‌توان از فایل‌های ثبت وقایع در سمت سرور نیز استفاده کرد [2][3]. پس از آن می‌توان با داشتن سابقه رفتار کاربر در مواجهه با نتایج بدست آمده با استفاده از الگوریتم‌های هوشمندی نظیر الگوریتم ژنتیک و یا الگوریتم ازدحام ذرات و یا نسل‌های جدیدتر این الگوریتم‌ها و ترکیب خوشه‌های داده و سابقه کاربر در موارد جستجو بعدی نتایج مورد نظر کاربر را کشف کرد و آنها را مقدم بر سایر نتایج در بهترین نقطه صفحه و در اولین پیشنهادات قرار داد تا کاربر با احتمال بیشتر و صرف هزینه کمتر به جواب مورد نظر خویش برسد. با توجه مطالب بیان شده، خوشه‌بندی یک کلاس‌بندی بدون نظارت است که در آن کلاس‌ها از پیش تعریف نشده‌اند. خوشه‌بندی داده‌ها موضوع مهمی در زمینه کاوش اطلاعات محسوب می‌شود و به عنوان یکی از فازهای پیش پردازش داده عمل کرده و نتایج ارزشمندی را در پی دارد. خوشه‌بندی بر روی وب اهمیت فراوانی دارد. زیرا، علاوه بر پیش پردازش، می‌تواند برای بهبود نتایج موتورهای جستجو و یا بهبود عملیات خزشی وب به کار رود. با این کار علاوه بر این که نیاز به سازماندهی دستی اطلاعات برطرف می‌شود، محدود کردن جستجو به تعداد

محدودی خوشه، می‌تواند کارآیی بازیابی را بهبود بخشد. در نهایت این کار اجازه می‌دهد تا مجموعه مستندات آسانتر توسط کاربران مشاهده شوند. از لحاظ کاربرد خوشه‌بندی در وب، کارهای انجام شده می‌تواند در دو دسته قرار گیرد: خوشه‌بندی مستندات وب در دامنه خاص و خوشه‌بندی نتایج جستجو. خوشه‌بندی مستندات وب چالش‌های متعددی را موجب می‌شوند که برخی از آنها شامل مشخص کردن مشخصه‌های مربوط به مستندات و تعیین وزن مناسب برای هر یک از آنها، انتخاب یک روش خوشه‌بندی و برقراری یک معیار مناسب شباهت و همین‌طور محدودیت‌های منابع محاسباتی و حافظه می‌باشد. بنابراین، هدف از انجام خوشه‌بندی ارائه چشم انداز مناسبی از اتفاقات در حال وقوع در پایگاه داده‌ها به مصرف‌کننده نهایی اطلاعات می‌باشد. کاربرد دیگر خوشه‌بندی را می‌توان در تعیین داده‌هایی که با سایر داده‌ها تفاوت چشمگیر دارند عنوان نمود. در خوشه‌بندی کلاسیک، هر نمونه ورودی متعلق به یک و فقط یک خوشه است و نمی‌تواند عضو دو خوشه و یا بیشتر باشد. به عبارتی خوشه‌ها همپوشانی ندارند در حالیکه در خوشه بندی فازی یک نمونه می‌تواند متعلق به بیش از یک خوشه باشد. خوشه‌بندی فازی به کشف مدل های فازی از داده‌ها می‌پردازد. این سمینار به بررسی الگوریتم‌هایی که هدف خوشه‌های فازی کاربران وب می‌باشد می‌پردازد.