ISSN: 2745-3673
http://www.Rcsj.ir

مجلـه‌ی علمی پژوهش در علوم رایانه (سال پنجم)
شماره ۱۷ / بهار ۱۳۹۹ / ص ۱۲-۱

ارائه یک سیستم توصیه گر برای استخراج اطلاعات مورد نیاز کاربران از کتابخانه های دیجیتالی بزرگ

رضا مولایی فرد
کارشناسی ارشد کامپیوترنرم افزار

نام نویسنده مسئول:
رضا مولایی فرد

تاریخ دریافت: ۱۳۹۸/۱۱/۲۳
تاریخ پذیرش: ۱۳۹۹/۱/۱۹

چکیده

امروزه کتابخانه های دیجیتال یکی از بهترین راهکارها برای به دست آوردن اطلاعات مورد نیاز کاربران می باشد. این کتابخانه های دیجتال با توجه به دسترسی آسان تر و هزینه کمتر یکی از بهترین راهکارهای کاربران برای بدست آوردن کتاب های موردنظر کاربران می باشد ولی با توجه به رشد روز افزون تعداد کتاب های مختلف در زمینه های مختلف وجود سیتمی که بتواند اطلاعات مورد نیاز افراد را از میان حجم عظیم این کتاب ها که روز به روز در حال افزایش می باشند، استخراج کند لازم و ضروری به نظر می رسد،یکی از بهترین راهکارها برای بدست آوردن کتب موردعلاقه کاربران از میان حجم عظیم داده ها، شخصی سازی این سیستم ها می باشد. یکی از بهترین سیستم های شخصی سازی استفاده از سیستم های توصیه گر می باشند، سیتم های توصیه گر یا پیشنهاد دهنده سیستم هایی هستند که با یکسری از روش های داده کاوی و روش هایی مانند اینکه کاربر در گذشته چه کتاب هایی جستجو کرده است می تواند پیشنهاد های مناسبی به کاربر ارائه دهند.در این تحقیق از یک سیستم توصیه گر مبتنی بر پالایش مشارکتی استفاده شده است که با دسته بندی کابران مشابه در یک گروه و با توجه به سوابق این افراد و امتیازهایی که به داده ها در پروفایل خود داده اند و تشابهات بین آنها به کاربر موردنظر می تواند پیشنهاد مناسبی از بین این حجم کتب مختلف ارئه دهد که موردنظر و مورد علاقه کاربر باشد.

واژگان کلیدی: سیستم توصیه گر، استخراج اطلاعات، داده کاوی، کتابخانه دیجیتالی، فیترینگ مشارکتی.

ISSN: 2745-3673
http://www.Rcsj.ir

مقدمه

با توجه به رشد روز افزون داده ها و کتب و همچنین با توجه به مشکلاتی که جهت تهیه کتب مختلف وجود دارد، وجود یک سیستم الکترونیکی که بتواند مجموعه ایی از این کتب را در اختیار کاربران قرار داده و مشکلات دسترسی به کتب را حل کند لازم و ضروری به نظر می رسد.کتابخانه های دیجیتال راه حل مناسبی برای دسترسی به کتب مورد نیاز افراد می باشد. کتابخانه های دیجیتالی یکی از زمینه های پژوهشی است که امروزه در برخی از رشته ها دنبال می شود. کوشا در مقاله خود تعاریف مختلف را که نتیجه چهار دیدگاه مخلف در زمینه کتابخانه هستند شناسایی نموده که عبارتند از؛ تعاریف ارائه شده توسط متخصصان رایانه: در این دیدگاه کتابخانه دیجیتال در قالب پایگاه اطلاعاتی و نظام بازیابی اطلاعات تعریف می شود: تعاریف ارائه شده توسط محققان کتابداری و اطلاع رسانی که در آن کتابخانه دیجیتال به مثابه سازمان و خدمات است؛ تعرف ارائه شده توسط سازمان های مجری و سیاست گذار که به دو دسته تعاریف ارائه شده از طرف سازمان های سیاست گذار کتابخانه های دیجیتال مانند فدراسیون بین المللی کتابخانه های دیجیتال و تعاریف ارائه شده توسط سازمان های مجری کتابخانه های دیجیتال تقسیم می شوند؛ و تعاریف پژوهش مدار که در آن ها کتابخانه دیجیتال به منزله یک مسئله پژوهشی در نظر گرفته و تعریف می شوند. جوامع مختلفی در امر کتابخانه دیجیتال دخیلند و هرکدام رویدادها، مفاهیم، معانی، تمرکز و رویکرد خود را دارند. دو گروه عمده که به پژوهش در زمینه کتابخانه های دیجیتال می پردازند عبارتند از علم و مهندسی کامپیوتر و علم و اطلاعات و دانش شناسی. برخی از موضوعات مورد توجه در عرصه علوم کامپیوتر عبارتند از: تعامل بین مجموعه های متجانس، یکپارچه سازی داده ها مانند ویدیو،متن،صوت،استانداردها و پروتکل شبکه ایی، موتورهای جستجو و عامل های جستجو، محدودسازی، تلخیص و یکپارچ سازی داده ها را تسهیل می کنند، فناوری های مصورسازی و تعاملی مانند امکان مرور حجم وسیعی از متون یا تصاویر. رویکردهای غالب در علم اطلاعات و دانش شناسی نیز عبارتند از: رویکرد سازمان، خدمات و فعالیت محور: در این رویکرد مفهوم کتابخانه گسترده می شود؛ رویکرد محتوا و مجموعه محور: که در آن منابع اطلاعاتی مختلف و همچنین مخازن مختلف منابع دیجیتال موردتوجه قرار می گیرند؛ رویکرد دسترسی به مجموعه ها: که در این رویکرد به مسائلی نظیر محیط ها و جوامع کاربردی مختلف و میزان یکپارچه سازی یا جداسازی منابع اطلاعاتی می پردازد (saracevic 1999). با وجود همه این تقاسیر یک راه حل مناسب جهت استخراج اطلاعات،استفاده از سیستم های توصیه گر می باشند، سیستم های توصیه گر یا پیشنهاد دهنده سیستم هایی هستند که با یکسری از روش های داده کاوی می توانند پیشنهادات مناسبی را به کاربر برای استخرج طلاعات ارائه دهند.در این تحقیق از یمک سیستم توصیه گر مبتنی برفیلترینگ مشارکتی استفاده می شود که می تواند اطلاعات مورد نیاز کاربر را استخراج و پیشنهادات مناسبی را به وی ارائه دهند که موردنظر کاربر باشند.

پیشینه پژوهش

کیم و لی در سال ۲۰۰۸ در ژژوهشی با عنوان بررسی ساختار درونی مطالعات آرشیوی با روش متن کاوی در سال های ۲۰۰۱ تا ۲۰۰۴،به بررسی ۴۳۲ مقاله از سال ۲۰۰۱ تا ۲۰۰۴ پرداخته و ۴۳ خوشه از اسناد با استفاده از میانگین درون گروهی در نرم افزار SPSS ایجاد کرده اند. سپس شبکه های مسیریاب این ۴۳ خوشه را ساخته و در ۷ شاخه موضوعی شامل:کتابخانه های دیجیتال و فناوری آرشیوسازی دیجیتال، منابع آنلاین و کمک های اکتشافی، آرشیوها و آرشیویست ها،موضوعات سیاسی و حقوقی، مسائل فنی و پیشینه های الکترونیک، مدیریت اطلاعات و رکوردها و پست الکترونیک و اطلاعات حرفه ایی گروه بندی کرده اند.در نهایت این ۷ موضوع در ۳ بحش ادغام شده اند که شامل کتابخانه های دیجیتال، آرشیوها و پژوهش ها در عمل می شوند.این مطالعه تغییرات پویای موضوعات پژوهشی این رشته،از حوزه های موضوعی صرفا سنتی تا ظهور حوزه های موضوعی پیچیده را از سال ۲۰۰۱ تا ۲۰۰۴ نشان می دهد.نتایج این پژوهش بیان می کند که حوزه های پژوهشی در علوم آرشیوی پتانسیل رشد بالایی داشته وهمچنان توسعه خواهند یافت.

عرب در سال ۱۳۹۲ در مقاله ایی با عنوان، ارزیابی قابلیت های جستجو پنج گانه کتابخانه دیجیتالی ایرانی، به بررسی کتابخانه های دیجیتالی ایرانی (نور،دید،تبیان،میراث فرهنگی، صنایع دستی و گردشگری و ارم) پرداخته است. نتایج پژوش نشان

ISSN: 2745-3673
http://www.Rcsj.ir

داده در کتابخانه های دیجیتالی مذکور، هیچکدام به دو ویژگی جستجو، جستجوی مجاورتی و جستجوی کوتاه سازی توجه نداشته اند، ولی همه کتابخانه ها دو مورد جستجوی ساده و جستجوی پیشرفته را دارا می باشند.

زائو و زنگ در سال ۲۰۱۱ در مقاله ایی با عنوان،ترسیم نقشه علمی پژوهش های تولید شده چینی در حوزه کتابخانه در سال های ۱۹۹۴ تا ۲۰۱۰ برای دست یابی به پارادایم های فکری در حوزه کتابخانه های دجیتالی،با استفاده از روش های متن کاوی و تحلیل هم رخدادی واژگان به بررسی ۱۲۵۰ و ۶۰۶۸ مقاله بازیابی شده در حوزه کتابخانه دیجیتالی به ترتیب در پایگاه ملی دانش زیربنایی چین و پایگاه سانس دایرکت پرداختند .در نهایت آنها با استفاده از فنون تحلیل شبکه های اجتماعی،نقشه های موضوعی و الگوهای پژوهشی کتابخانه های دیجیتال در چین را با استفاده از نرم افزارهای UCINET و Netdraw ترسیم کردند.

ساروش پاریک در سال ۲۰۱۳ در مقاله خود به صورت کتاب سنجی به مطالعه متون مجلات ایفلا مربوط به سال های ۲۰۰۱ تا ۲۰۱۰ پرداخته و هدف از مطالعه خود را تحلیل کتاب سنجی از ویژگی هایی مانند میزان انتشار مقاله در سال،الگوهای نویسندگی،میزان همکاری موسسات،توزیع موضوعات،الگوی استناد و... بیان کرده است.

بالاجی بابو و کریشنامورسی در سال ۲۰۱۳ در پژوهشی با هدف مطالعه ی تجزیه و تحلیل الگویی از اتوماسیون کتابخانه برای کشف منابع بوسیله کاوش برنامه های کاربردی کشف منابع به این نتیجه رسید که،رشد صنعت اتوماسیون هند پر رونق است.با این حال سازگاری نرم افزار کتابخانه و پیشرفت جامعه رضایت بخش نیستند.

نوروزی در سال ۱۳۸۹ در پژوهشی تحت عنوان، بررسی میزان رعایت معیارهای ارزیابی رابط کاربر در صفحات وب فارسی کتابخانه های دیجیتالی خود ساخته و خریداری شده در ایران،به ارزیابی رابط کاربر کتابخانه های دیجیتالی ایران پرداخته است.روش انجام پژوهش از نوع تحقیقات ارزیابانه است که از یک سیاهه وارسی متشکل از ۱۰ معیار اصای و ۱۱۴ مولفه فرعی برای ارزیابی و تحلیل ابعاد پژوهش استفاده شده است.یکی از معیارهای اصلی این پژوهش،جستجو می باشد.این پژوهش نشان می دهد که از حمه امتیازات مربوط به معیار ده گانه،کتابخانه های دیجیتالی خودساخته توانستند در هفت معیار تصیح خطا،کنترل کاربر،راهبری،جستجو وسادگی امتیاز بالایی را کسب کنند.در صورتی که کتابخانه های دیجیتالی خریداری شده،تنها در سه معیار زبان رابط کاربر،نمایش اطلاعات و انسجام وضعیت بهتری قرار دارند.

سیستم توصیه گر

سیستم های توصیه گر[1] یا پیشنهاد دهنده زیر مجموعه ایی از سامانه پالایش اطلاعات که به دنبال پیش بینی امتیاز یا اولویتی است که کاربر به یک آیتم (داده،اطلاعات،کالا و...) خواهد داد. در سال های اخیر سیستم های توصیه گر بسیار متداول شده و در حوزه های مختلفی مورد استفاده قرار گرفته اند. برخی از کاربردهای معروف این سیستم ها می توان به موارد زیر اشاره کرد: موسیقی، صفحات وب، اخبار، کتاب ها و مقالات، صنعت گردشگری، صنعت هتل داری، صنعنت پزشکی، جست و جو و شبکه های اجتماعی. در واقع سامانه های توصیه گر مانند یک فیلتر عمل می کنند، فیلتری که فقظط آنچه مطلوب کاربر است را به او نشان می دهد که به این عمل شخصی سازی کردن اطلاعات می گویند. به طور کل سامانه توصیه گر یک سامانه پشتیبان شخصی سازی است که اطلاعات را به سه عامل تعیین کننده، سفارشی سازی، علاقه مندی و سودمندی، برای کاربران ویژه بوسیله تجزیه و تحلیل سلایق آنها و محتوای آیتم ها می سنجد. سیستم توصیه گر از جمله ابزارهایی است که می تواند کاربران را در محیط های الکترونیکی به سمت یافتن اطلاعات، خدمات و آیتم های مورد نظرشان هدایت کند [کونستانت و همکاران، ۲۰۰۸]. سیستم های توصیه گر با قابلیت کشف علایق کاربران و پیش بینی اولویت آنها،آیتم هایی که احتمال می رود مورد توجه کاربر باشد را از بین حجم بالای داده ها پالایش کرده و یا آنها را پیشنهاد آنها، در زمان او صرفه جویی می کند. از طرف دیگر این سیستم ها توانایی ذخیره و تحلیل رفتارهای گذشته کاربر،خدمات و اطلاعاتی را که مورد توجه کاربران نبوده و احتمالا به آنها علاقه مند هستند را نیز استنتاج کرده و نتایج جالب توجهی به کاربران ارائه می کند.در واقع سیستم های توصیه گر یکی از ابزارهای اصلی غلبه بر مشکل افزونگی اطلاعات بوده و با داشتن قدرت تحلیل رفتارهای کاربر، مکمل هوشمندی

[1] Recommender system

ISSN: 2745-3673
http://www.Rcsj.ir

برای مفاهیم بازیابی و پالایش اطلاعات است.امروزه سیستم های توصیه گر در زمینه های متنوعی از پالایش اطلاعات موجود در وب متناسب با خواسته های کاربر تا تجارت الکترونیکی،پیشنهاد فیلم،موزیک،کتاب،مقاله و... کاربرد دارند.

سیستم های توصیه گر انواع مختلفی دارند که به طور کلی به دسته های زیر تقسیم می شوند:

سیستم توصیه گر مبتنی بر محتوا

این روش بر مبنای توصیف آیتم ها با کلمات کلیدی و ایجاد پروفایل براساس اولویت های کاربر است. در واقع، این الگوریتم مواردی را براساس تشابهات با علایق کاربر در گذشته توصیه می کند، یعنی مواردی که قبلا توسط کاربر امتیازدهی شده با آیتم موجود مقایسه کرده و بهترین تطابق را توصیه می کند. این سیستم با استفاده از مدل اولویت کاربر وسابقه تعامل کاربر پروفایلی را ایجاد می کند. این سیستم همچنین پروفایلی مبتنی بر محتوا براساس بردار وزن ویژگی ها ایجاد می کند. این وزن ها دلالت بر اهمیت ویژگی ها توسط کاربر دارند و بر اساس عقیده کاربر افزایش یا کاهش می یابند (رائو کاگیتا و همکاران، ۲۰۱۵؛ لو وهمکاران، ۲۰۱۵؛ وانگ و همکاران ۲۰۱۵).

سیستم توصیه گر مبتنی بر فیلترینگ مشارکتی

مبنای فیلترینگ مشارکتی بر اولویت های رفتاری یا فعالیت های کاربران و پیش بینی علایق آنان، براساس تشابهات با کاربران دیگر است. این روش با پیش بینی خودکار فیلترینگ،علایق کاربران، اطلاعات را جمع آوری می نمایند. این روش معمولا به مشارکت کاربر، یعنی تجزیه و تحلیل پروفایل و الگوریتمی برای یافتن افراد با علایق مشابه نیاز دارد. در واقع، در این روش کاربر نظر خود را با امتیاز دهی اقلام در سیستم بیان می کند و سیتم های کاربرانی را با الگوهای امتیاز دهی یکسانی را به اشتراک می گذارند یافته و از این کاربران همفکر برای محاسبه پیش بینی استفاده می کند (پاول و همکاران،۲۰۱۵؛کیروش داسیلوا وهمکاران،۲۰۱۶؛یانگ و همکاران ،۲۰۱۴؛وی وهمکاران،۲۰۱۵).

سیستم توصیه گر مبتنی بر دانش

این سیستم ها بر اساس ادراکی که از نیاز های کاربر و ویژگی های آیتم ها پیدا کرده اند، توصیه ارائه می دهند. به عبارتی در این گونه از سیستم های توصیه گر مواد اولیه مورداستفاده برای تولید لیستی از توصیه ها، دانش سیستم در مورد کاربر و آیتم است [بورکه،۲۰۰۰].سیستم های مبتنی بر دانش از متد های مختلفی که برای تحلیل دانش، قابل استفاده هستد بهره میبرند که متد های رایج در الگوریتم های ژنتیک،فازی،شبکه های عصبی و...از جمله آنها. همچنین در این گونه سیستم ها از درخت ای تصمیم، استدلال های مبتی بر شاهد و... نیز می توان استفاده کرد. یکی از رایج ترین متدهای مبتنی بر دانش در سیستم های توصیه گر، روش استدلال مبتنی بر نمونه است.

سیستم های توصیه گر ترکیبی

در سیستم های توصیه گر ترکیبی برای رسیدن به بالاترین کارایی، بر اساس یک استراتژی معین تکنیک های مختلف با یکدیگر ترکیب می شوند [بورکه،۲۰۰۲]. برای مثال دو تکنیک پالایش مشارکتی و متنی بر دانش در صورتی که با یکدیگر ترکیب شوند، نتیجه سیستمی خواهد بود که به واسطه جزء مبتنی بر دانش، می تواند مشکل شروع آهسته تکنیک پالایش مشارکتی را جبران کند و از طرفی با وجود جزء پالایش مشارکتی و قدرت بالای آن در یافتن اولویت های مشابه کاربران می تواند توصیه ها را تولید کند که هیچ سیستم مبتنی بر دانش، توانایی توصیه به آن را بر اساس دانش ندارد.

ISSN: 2745-3673
http://www.Rcsj.ir

داده کاوی

مرکز تحقیقات آمریکا و اداره پاسخگویی سازمان ها داده کاوی را مستلزم استفاده از ابزارهای پیشرفته برای تحلیل و کشف روابط و استخراج الگوهای ارزشمند داده ها به منظور دسترسی به قوانین جدید معنا دار می داند (مرادی و قاسمی؛ ۱۳۹۷). کشف دانش درون داده ها در عصر اطلاعات یکی از هیجان انگیزترین و کلیدی ترین مفاهیمی است که روز به روز به اهمیت آن افزوده می شود (رحمانی و حاجی زین العابدینی؛ ۱۳۹۵). با توجه به تعاریف و تفاسیر مطرح شده از دیدگاه های مختلف، می توان دو جزء اساسی را در داده کاوی مشخص نمود، اولی کشف الگوهای پنهان در داده ها می باشد و دیگری استفاده از این الگوها برای پیش بینی نتایج در آینده است (مرادی و قاسمی؛۱۳۹۷). در سال های اخیر با پیشرفت سریع و گسترده شبکه ها و فناوری پایگاه های اطلاعاتی و کتابخانه ها هم با حجم بسیار زیادی از اطلاعات مواجه بوده اند. با گسترش کارایی کتابخانه های دیجیتالی و انبوه منابع اطلاعاتی ناهمگن در حال رشد که جان نثبرت، معمای، غنای اطلاعاتی و فقر دانش را مطرح می نماید (پناهی؛ ۱۳۹۶) در آن فناوری داده کاوی با دو اصل مهم پیش بینی نتایج و کشف دانش در اثرسنجی وب جهان گستر در این گونه کتابخانه ها و مهم تر در جذب کاربران خاص با منابع اطلاعاتی متنوع ارزش کاربردی بسیاری پیدا کرده است.

وب کاوی

عبارت وب کاوی مترادف با یکی از عبارت های استخراج دانش،برداشت اطلاعات، وارسی داده ها و حتی لایروبی داده هاست که در حقیقت کشف دانش در پایگاه داده ها را توصیف می کند. بنابراین ایده ایی که مبنای داده کاوی است یک فرایند با اهمیت از شناخت الگوهای بالقوه مفید و قابل درک داده هاست. داده کاوی یا به تعبیر دیگر کشف دانش در پایگاه داده ها، استخراج بدیهی اطلاعات بالقوه مفید از روی داده هایی است که قبلا ناشناخته مانده اند.این مطلب برخی از روش های فنی مانند خوشه بندی، خلاصه سازی داده هاؤفراگیری قاعده های رده بندی، یافتن ارتباط شبکه ها، تحلیل تغییرات و کشف بی قاعدگی را شامل می شود (سعیدی ،۱۳۸۴).

وب کاوی اشاره به کلیه فعالیت های داده کاوی و فنون وابسته دارد که برای کشف خودکار و استخراج دانش از اسناد و خدمات وب به کار می رود (ول و رویاکرز،۲۰۰۴). اطلاعات بسیار زیاد و ناهمگونی در محیط وب وجود دارد که که باعث می شود کسب دانش موجود در محتوای صفحات وب مشکل تر می شود،بنابراین در چنین محیطی به کارگیری ابزارها و فنون داده کاوی برای کشف اطلاعات و دانش مرتبط ضروری است (سالاری:۱۳۸۳).

انواع روش های وب کاوی

وب کاوی ممکن است به سه شیوه انجام گیرد؛کاوش محتوای وب،کاوش ساختار وب و کاوش کاربری وب می باشد.

کاوش محتویات وب؛فرایندی است که برای بدست آوردن مدل یا دانش ارزشمند و بالقوه از محتویات اسناد، اطلاعات توصیفی و یا نتایج جستجو در صفحات وب که هم زمان می تواند برای بدست آوردن اطلاعات مفید از ساختار وب و ارتباطات پیوندها به کار گرفته می شود.

کاوش ساختار وب؛ فرایندی است که ارتباطات پیوندهای بین مسیر صفحات، ساختار اسناد و ساختار مسیر در آدرس اینترنتی را جستجو می کند. در فضای وب علاوه بر محتویات صفحات،ساختار آن ها نیز اطلاعات مفیدی دارد، اگر ارجاعات به پیوندی افزایش یابد،پس آن صفحه مهم است و باید مسیر جستجو را بر اساس آن تغییر داد.

کاوش کاربری وب: فرایندی است که الگوی دسترسی به صفحات را با بررسی لاگ و داده های مربوط به آن کشف می کند. لاگ های سروری که همان اطلاعات تولید شده در هر ارتباط کاربر و سرور در جهان ثبت می شود.با بررسی این اطلاعات می توان رفتار کاربر را شناخت و ساختار صفحه وب را با هدف شخصی سازی بهبود بخشید.[ژائو و همکاران:۲۰۱۴].

ISSN: 2745-3673
http://www.Rcsj.ir

خوشه بندی

خوشه بندی یکی از بهترین روش هایی است که برای مدل سازی داده ها ارائه شده است و اطاعات را بر اساس شباهت به خوشه ها تقسیم می نماید. قابلیت آن در ورود به فضای داده و تشخیص ساختار آن ها، خوشه بندی را یکی از ایده آل ترین مکانیزم ها برای کار با دنیای عظیم داده ها کرده است.در خوشه بندی بدون اینکه هیچ مدل از پیش معینی داشته باشیم به دنبال یافتن مدل های مشترکی هستیم که د داده ها وجود دارد.ایده ی این روش د دهه ۱۹۳۵ ارائه شد و در حال حاظر با پیشرفت ها و جهش های عظیمی که در آن پدید آمده است کاربردهای مختلفی پیدا کرده است.

در سیستم های توصیه گر می توان از خوشه بندی برای دسته بندی کردن داده های مشابه با هم استفاده کرد تا جستجو و ساخت پیشنهادات فقط روی خوشه هایی انجام شود که پیش بینی می شود برای کاربر جالب تر خواهد بود. در نتیجه از تکنیک های خوشه بندی برای افزایش مقیاس پذیری الگوریتم های پشتبان سیستم های توصیه گر استفاده شده است. این مساله به خصوص در شرایطی که برای پیشنهاد آیتمی لازم است تمام فضای مساله جستجو شود در کارایی الوریتم تاثیرات مثبتی نشان داده است [نکیتیم؛۲۰۰۹].

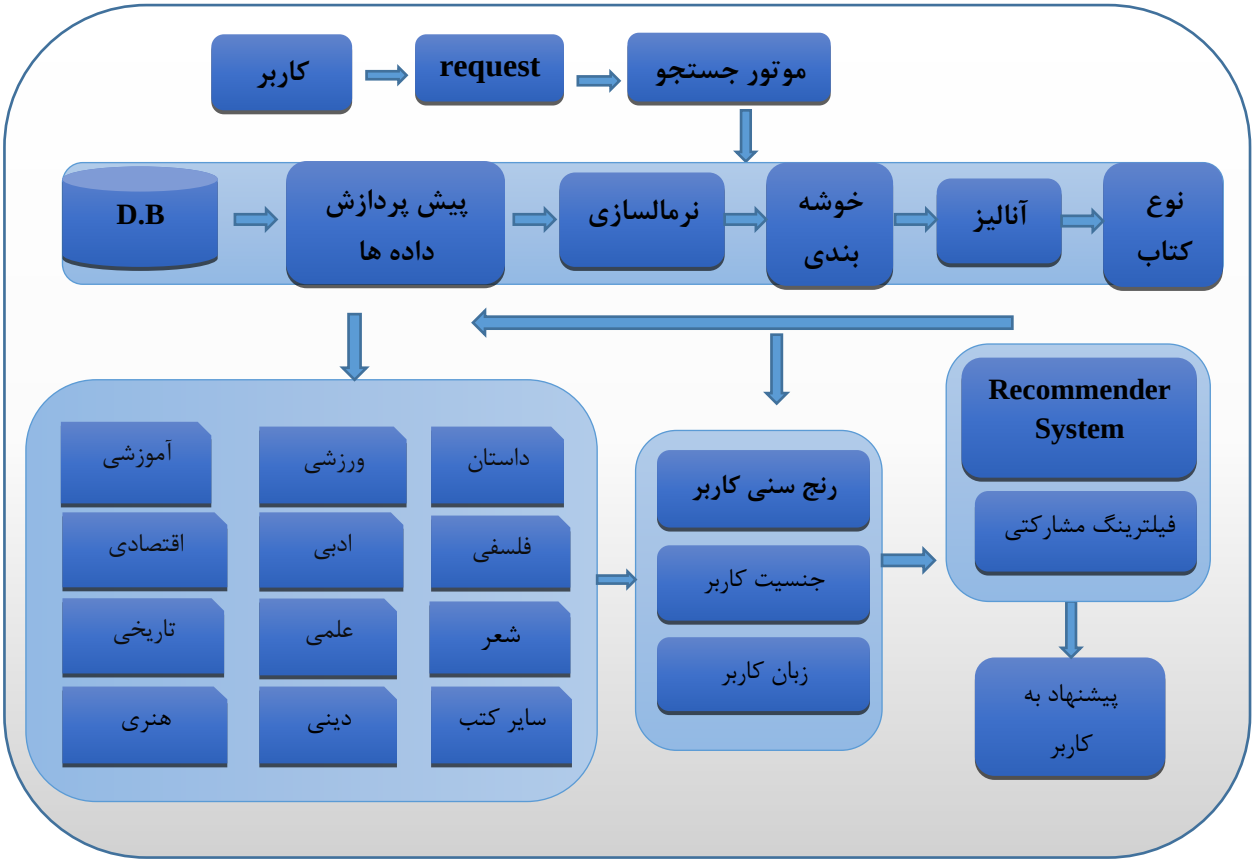
پروفایل کاربر

در سال های اخیر تعداد زیادی سیستم توصیه گر برای پیش بینی علایق و ارائه توصیه به کاربران،طراحی و پیاده سازی شده است. در هر یک از این سیستم ها با توجه به حوزه کاری و اهداف، مجموعه ایی از تکنیک های ساخت، به روز رسانی و استخراج داده ها به کار گرفته شده است ولی محور اساسی در تمامی این سیستم ها پروفایل کاربر است. چگونگی ساخت پروفایلی که در ساخت توصیه ها استفاده خواهد شد، پروفایل پیش فرض سیستم برای کاربران، نحوه بروز رسانی اطلاعات پروفایل و منبع این بروز رسانی فاکتورهایی هستند که در طراحی یک سیستم توصیه گر جایگاه مهمی دارند [مونتانار و همکاران؛ ۲۰۰۳].

روش پیشنهادی موردنظر

در روش پیشنهادی موردنظر پس از جمع آوری پایگاه داده ایی از کتب مختلف ابتدا عمل پیش پردازش داده ها را برروی Data انجام می دهیم سپس باید Data را خوشه بندی کنیم برای این کار از روش خوشه بندی K-means استفاده می کنیم. پس از خوشه بندی Data باید داده ها را براساس نوع کتاب ها به دسته های مختلف (آموزشی، تاریخی، درسی، اقتصادی، ورزشی، سیاسی و...)دسته بندی می کنیم.سپس با اطلاعاتی که کاربر در پروفایل خود وارد می نماید، سیستم توصیه گر مبتنی بر فیلترینگ مشارکتی براساس داده های موجود و آنالیزهای انجام شده بهترین تصیم گیری را انجام می دهد و بهترین پیشنهاد را به کاربر ارائه می دهد که این پیشنهادات می توانند موردعلاقه کاربر باشند.در ادامه به شرح هریک از مراحل این روش پرداخته می شود.در شکل(۱)نمایی از روش پیشنهادی را مشاهده می کنید.

ISSN: 2745-3673
http://www.Rcsj.ir



شکل(۱)چارچوب پیشنهادی روش موردنظر

پیش پردازش داده ها

در مرحله اول روش پیشنهادی ابتدا باید عمل پیش پردازش داده ها را انجام دهیم در فرایند داده کاوی مانند طبقه بندی و خوشه بندی نیاز داریم تا داده ها برای الگوریتم آماده شوند،زیرا معمولا نمی توان داده ها را به صورت خام به الگوریتم های داده کاوی و یادگری ماشین تزریق کرد.برای آماده سازی داده ها،نیاز است تا آنها را از شکل و حالت اولیه،خارج کرده و به شکلی که برای الگوریتم مناسب باشد تبدیل کرد.همچنین داده های موجود معمولا دارای زواید مختلفی هستند که ممکن است الگوریتم را دچار خطا کنند.در داده کاوی هم نیاز داریم تا داده های اضافی که به مسئله و الگوریم کمکی نمی کنند را حذف کنیم.عملیات پیش پرازش داه ها معمولا قبل از عملیات اصلی الگوریتم های داده کاوی انجام می گیرند و باعث تسهیل و کمک به الگوریتم ها می شوند.پیش پردازش داده ها گامی مهم در راستای داده کاوی موفقیت آمیز است.اعمالی که در آماده سازی داده ها انجام می شود عبارتند از پاکسازی داده ها ،یکپارچه سازی داده ها،تبدیل داده ها و کاهش داده ها می باشد.بر اساس نوع کاربردی که عمل داده کاوی باید روی آن انجام شود،تکنیک های مختلفی برای هر یک از این اعمال صورت می گیرد.

نرمال سازی داده ها

پس از پیش پردازش داه ها باید داده ها را نرمال کنیم،نرمال سازی داده ها تغییر داده ها به گونه ایی است که آنها را به یک دامنه کوچک و معین مانند فاصله بین ۱- و ۱ نگاشت کنند.هدف نرمال سازی حذف افزونگی داد و باقی نگهداشتن وابستگی بین داده های مرتبط می باشد.این فرایند اغلب باعث ایجاد جداول بیشتر میشود ولی اندازه گیری پایگاه داده را کاهش داده و بهبود کارایی را تضمین می کندروش های مختلفی جهت نرمال سازی داده ها وجود دارد که معروفترین آنها می توان به

ISSN: 2745-3673
http://www.Rcsj.ir

روش MINMAXNORMALIZATINاشاره کرد. در این روش هرکدام از داده ها را می توان به یک بازه دلخواه تبدیل کرد. فرمول کلی این روش برای تبدیل داده ها به بازه ی بین ۱تا۰ به صورت رابطه (۱) می باشد:

رابطه(۱)

$$Z = \frac{X - MIN(X)}{MAX(X) - MIN}$$

خوشه بندی داده ها

در این تحقیق از الگوریتم k-means به عنوان الگوریتم خوشه بندی استفاده می شود،این الگوریتم یکی از رایج ترین الگوریتم های استفاده شده در سیستم های توصیه گر می باشد.به عنوان معیار شباهت نیز ضریب همبستگی پیرسون استفاده می شود.در واقع در این بخش تعداد داده ها خوشه بندی می شود و برای هر خوشه یک نماینده و یا یک بردار میانگین در نظر گرفته می شود بدین صورت که تعداد کتب دیجیتالی به عنوان ورودی به الگوریتم خوشه بندی داده می شود و الگوریتم با استفاده ضریب همبستگی پیرسون کتاب ها را به خوشه هایی تقسیم می کند تا در هرخوشه بیشترین شباهت و در بین خوشه ها کمترین شباهت وجود داشته باشد. ضریب همبستگی پیرسون به صورت رابطه(۲)محاسبه می شود:

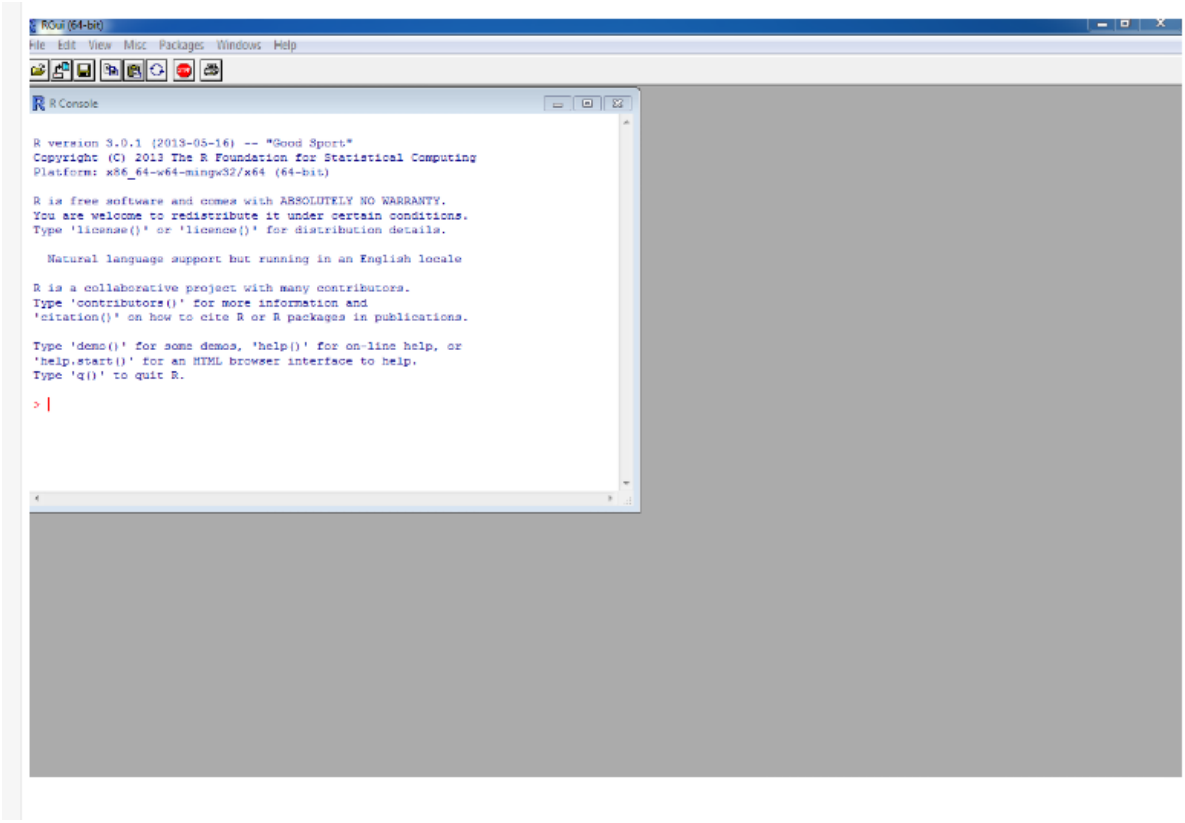
رابطه(۲)

$$SC(u,v) = \frac{\sum_j^m (r_{u,j} - \bar{r}_u)(r_{v,j} - \bar{r}_v)}{\sqrt{\sum_j^m (r_{u,j} - \bar{r}_u)^2 (r_{v,j} - \bar{r}_v)^2}}$$

آنالیز Data

پس از خوشه بندی داده ها نیاز داریم تا آنها را بر اساس انواع کتاب های مختلف آنالیز کنیم.نرم افزارهای زیادی برای آنالیز اطلاعات خوشه بندی شده موجود می باشد،در این تحقیق برای آنالیز داده های خوشه بندی شده از نرم افزار R استفاده می کنیم و توسط آن کتاب ها را به دسته بندی های مختلف دسته بندی می کنیم.

نمایی از برنامه R را در شکل زیر مشاهده می کنید:



مجلـه‌ی علمی پژوهش در علوم رایانه ، شماره ۱۷، بهار ۱۳۹۹، ص ۱۲- ۱

ISSN: 2745-3673
http://www.Rcsj.ir

خصوصیات کاربر

در این مرحله با توجه به ویژگی هایی مانند رنج سنی کاربر،جنسیت کاربر،زبان کاربر و...سیستم توصیه گر سعی خواهد کرد که پیشنهادات خود را به کاربر ارائه دهد.

فیلترینگ مشارکتی

پالایش مشارکتی فرایند پالایش یا ارزیابی آیتم ها با استفاده از نظرات کاربران است.با وجود اینکه از شکل گیری اصطلاح پالایش مشارکتی تنها کمی بیش از یک دهه می گذرد،اما فلسفه این روش این است که استفاده از نظرات دیگران در تصمیم گیری است،برای قرن ها مورد استفاده ی انسان ها بوده است.برای مثال اگر دوستان شما از کتابی تعریف کنند راغب به خواندن آن کتاب خواهید شد.یا برعکس اگر کتابی را بد توصیف کنید بعید است برای خرید آن کتاب اقدام کنید.بعلاوه بعد از مدتی شما خواهید فهمید که نظرات کدام دوستانتان به نظرات شما نزدیکتر است و بتدریج فقط به نظرات آن دسته از دوستانتان که به شما شباهت دارند توجه خواهید کرد.کامپیوتر و بستر وب این امکان را فراهم کردند که قدم از ارتباطات شخصی فراتر گذاریم و به جای تصمیم گیری بر مبنای ده یا صدنفر از دوستان و نزدیکان،از نظرات بیش از میلیون ها کاربر استفاده کنیم.سرعت کامپیوتر این اجازه را می دهد که این نظرات را به صورت بلادرنگ پردازش کرده و بدانیم که افرادی که به ما شبیه هستند در مورد یک آیتم خاص که از آن بی اطلاعیم چه نظری دارند.

نحوه امتیاز دهی

در این قسمت به مفهوم امتیاز و ماتریس امتیاز پرداخته می شود.از آنجایی که در تکنیک پالایش مشارکتی نظرات کاربران نقش کلیدی دارد لازم است روش ها و قالب هایی برای جمع آوری آن طراحی شود.در ادبیات سیستم های توصیه گر روش های مختلفی برای جمع آوری نظرات کاربران معرفی شدند،اما روش معمولی که اکثر سیستم های توصیه گر مبتنی بر پالایش به کاربر گرفته می شود در نظر گرفتن یک بازه عددی(مثلا ۱تا۵)برای هر آیتم،تعریف مفهوم هر یک از این اعداد(مثلا۱:بسیار بد،۲:بد،۳:متوسط،۴:خوب،۵:بسیار خوب) و در خواست از کاربر برای نگاشت یکی از این اعداد به هر یک از آیتم هایی که مشاهده می کند.این اعداد در ادبیات سیستم های توصیه گر امتیاز خوانده می شود و این روش امتیازدهی نام دارد

در جدول زیر ماتریس امتیاز یک سیستم توصیه گر مبتنی بر فیلترینگ مشارکتی چند کتاب را مشاهده می کنید:

امتیاز دهی بدین صورت می باشد که به عنوان مثال کاربر ۳ به کتاب بیگانه امتیاز ۳ و کاربر ۱ به کتاب قلعه حیوانات امتیاز نداده است.

شناسه کاربر	کتاب قلعه حیوانات	کتاب رهایی از زندان ذهن	کتاب ۱۹۸۴	کتاب بیگانه
کاربر ۱		۵	۲	۳
کاربر ۲	۴		۳	
کاربر ۳			۵	۳
کاربر ۴	۵	۵		۴

با استفاده از سیستم توصیه گر مبتنی بر پالایش مشارکتی کتاب هایی را به کاربر توصیه می کند که می تواند موردعلاقه کاربر باشد.برای پیش بینی کتاب ها،از تکنیک های CF انجام می شود،تکنیک های CF از یک دیتابیس از ترجیحات برای آیتم های توسط کاربران استفاده می کنند که برای تخمین موضوعات اضافی یا ایجاد کاربران جدید است.در یک سناریوی معمولی CF،لیستی از m کاربر وجود دارد{u1,u2,u3..um}و همچنین یک لیست از n آیتم {i1,i2,i3...in} و هر کاربر Ui،یک لیست از آیتم ها دارد lui که کاربر آنها را رتبه بندی کرده است،و یا آنهایی که ترجیحات آنها از طریق رفتارشان استنباط کرده است.رتبه می تواند اشاره های صریح باشد و... که روی مقیاس ۱تا ۵ است و یا همچنین می تواند اشاره ضمنی باشد.

ISSN: 2745-3673
http://www.Rcsj.ir

الگوریتم نزدیکترین همسایه مبتنی بر کاربر

این نوع از الگوریتم ها در پیش بینی علاقه ی یک کاربر به یک آیتم خاص از امتیازهایی که کاربران مشابه او به آن آیتم داده اند استفاده می کند.این کاربران مشابه،اصطلاحا همسایگان کاربر نامیده می شوند.اگر n کاربر مشابه کاربر u باشد،گفته می شود که n همسایه u است.برای پش بینی میزان علاقه کاربر u به آیتم i باید میانگین امتیازهایی که همسایگان u(شامل کاربرn)به i داده اند را با تساوی(۱,۲)حساب کرد.در این تساوی r_{ni} امتیازی است که کاربر n به آیتم i داده است. نحوه محاسبه این الگوریتم را در رابطه (۲) مشاهده می کنیم.

رابطه(۳)

$$pred(u.i) = \frac{\sum \text{ncneighbors(u)r ni}}{\text{number of neighbors}}$$

پیشنهاد به کاربر

در مرحله آخر سیتم توصیه گر با تمام آنالیزهایی که انجام داده،لیستی از کتاب های مختلف را به کاربر پیشنهاد خواهد کرد که این کتاب ها می تواند مورد علاقه کاربر باشد.

نتیجه گیری

پیشرفت ها و تحولات پیش آمده در کتابخانه های دیجیتالی به عنوان نسل جدیدی از کتابخانه ها،ضروریت ارزیابی قابلیت های جستجو و نمایش نتایج جستجو را ایجاب می نماید،چرا که نظام جستجو و بازیابی هر کتابخانه دیجیتالی باید پاسخگوی تمام گروه های کاربری خود باشد،یعنی با استفاده از آن هم جستجوگران مبتدی و کم تجربه و هم چستجوگران با تجربه بتوانند به مطلوب خود برسند و با فراهم ساختن امکانات جستجوی متعدد اجازه انجام انواع مختلف جستجوها را به کاربران دارای سطوح مهارتی مختلف بدهند.در این تحقیق سعی شد با توجه به اینکه کتب مختلف روزانه در حال افزایش می باشند و دسرسی به این کتب با مشکلاتی از قبیل نحوه دسرسی،هزینه و...مواجه شده است به طراحی یک سیستم توصیه گر مبتنی بر پالایش مشارکتی برای استخراج کتب موردنیاز کاربران از کتابخانه دیجیتالی پرداخته شد که می تواند کتب موردنظر کاربر را از حجم وسیع اطلاعات موجود در کتابخانه دیجیتالی استخراج و بتواند پیشنهادات مناسب را به کاربر ارائه دهد.

ISSN: 2745-3673
http://www.Rcsj.ir

منابع و مراجع

[۱] پناهی، سمیه (۱۳۹۶). پژوهشی در باب خدمات شخصی سازی کتابخانه دیجیتال براساس داده کاوی. مجله الکترونیکی. متابدار۰/۲.دوره سوم.شماره اول (اردیبهشت ماه).

[۲] رحمانی ،محمود و حاجی زین العابدینی،محسن،(۱۳۹۵).کاربردهای داده کاوی در علم اطلاعات و دانش شناسی: نشریه مدیریت اطلاعات و دانش شناسی.مقاله ۲،دوره ۲،شماره ۳،پاییز ۱۳۹۴،صفحه ۲۳-۳۲.

[۳] سالاری،عبدالرضا.(۱۳۸۳) ”کاربرد محاسبات نرم در وب کاوی“،پایان نامه کارشناسی ارشد کامپیوتر، دانشگاه آزاد اسلامی وحد علوم و تحقیقات.

[۴] سعیدی،احمد.(۱۳۸۴)” داده کاوی،مفهوم وکاربرد آن در آموزش عالی“،شماره ۱۸،اسفند، ص ۲-۱۷.

[۵] کوشا، کیوان.۱۳۸۴.کتاب رقومی چیست؟ اصطلاحی رایج با مفهوم ابهام برانگیز. مطالعات کتابداری و سازماندهی، ۶۳:۹۷-۱۱۰.

[۶] مرادی،گلمراد؛قاسمی،وحید(۱۳۹۱)تکنیک داده کاوی و کاربرد آن در مطالعات اجتماعی.مجله علوم اجتماعی. بهار و نابستان، شماره ۱۱۹. (۱۷۶-۱۵۵).

[۷] نوروزی،یعقوب.۱۳۸۹.تحلیلی بر کاربر مداری رابط کاربر در صفحات وب فارسی کتابخانه های دیجیتالی ایران و ارائه پیشنهادی-فصلنامه علوم و فناوری اطلاعات،۱۶(۸):۷۹-۶۴.

[8] Babu, P.B. Krishnamurthy,M .2013. Library Automation to Resource Discovery: A Review of Emerging Challenges. Electronic Library, 31(4):433-451.

[9] Burke, R., ―Hybrid recommender systems: Survey and Experiments.‖ User Model. UserAdapt. Interact. 12, 4, 331–370, (2002)

[10] Kim, H. Lee, J. Y. 2008. Exploring the emerging intellectual structure of archival studies using text mining: 2001-2004. Journal of Information Science, 34(3): 356-369.

[11] Konstan, J. A., ―Introduction to recommender systems‖. In Proceedings of the 2008 ACM SIGMOD international Conference on Management of Data, Vancouver, Canada, (Jun, 2008).

[12] Lu Jie, Dianshuang Wu, Mingsong Mao, Wei Wang, Guangquan Zhang, (2015).Recommender System Application Developments: A Survey, Decision Support Systems, Volume 74,12-32.

[13] Montaner, M., Lopez, B., Rosa, and J.L.D.L.,‖taxonomy of recommender agents on the internet‖.Artificial Intelligence Review 19 (2003).

[14] Nkweteyim, D., ―Hyperlink Recommender Systems Design: a Research Study on Tools and Techniques.‖ VDM Verlag. (2009)

[15] Pareek, S. 2013. A Bibliometric analysis of the literature of IFLA Journal during 2001-2010. http://digitalcommons.unl.edu/libphilprac/954. (accessed 12/07/ 2013)

[16] Queiroz da Silva Edjalma, Celso G. Camilo-Junior, Luiz Mario L. Pascoal, Thierson C. Rosa, (2016). An evolutionary approach for combining results of recommender systems techniques based on collaborative filtering, Expert Systems with Applications,Volume 53, 204-218.

[17] R. Burke, ―Knowledge-Based Recommender Systems,‖ Encyclopediaof Library and Information Systems, A. Kent, ed., vol. 69,Supplement 32, Marcel Dekker, (2000).

[18] Rao Kagita Venkateswara, Arun K. Pujari, Vineet Padmanabhan,(2015). Virtual user approach for group recommender systems using precedence relations, Information Sciences, Volume 294, 15-30.

[19] Saracevic, T. 2000. Digital library evaluation: Toward evolution of concepts. Library trends, 49 (2):350-369. Shen, R. Gonçalves,M.A. Fox,E.A.. 2013. Key Issues Regarding Digital Libraries: Evaluation and Integration. Synthesis Lectures on Information Concepts, Retrieval, and Services, 5 (2):1-110. Tedd

[20] Wang, H., Li, W.-J.,(2014). Relational collaborative topic regression for recommender systems, IEEE Transactions on Knowledge and Data Engineering (TKDE), Volume 27, Issue 5,1343 - 1355.

ISSN: 2745-3673
http://www.Rcsj.ir

[21] Wang, H., Li, W.-J.,(2014). Relational collaborative topic regression for recommender systems, IEEE Transactions on Knowledge and Data Engineering (TKDE), Volume 27, Issue 5,1343 - 1355.
[22] Wei Jian, Jianhua He, Kai Chen, Yi Zhou, Zuoyin Tang, (2016). Collaborative Filtering and Deep Learning Based Recommendation System for Cold Start Items, Expert Systems with Applications, Volume 69, 29-39.
[23] Wel, litavan; Royakkers, Lamber(2004)"Ethical issue in web data mining". Ethics and Information Thecnology,6, pp. 129 140
[24] Yang Xiwang, Yang Guo, Yong Liu, Harald Steck,(2014). A survey of collaborative filtering based social recommender systems, Computer Communications, Volume 41,1-10.
[25] Zhao, L. Zhang, Q.2011. Mapping knowledge domains of Chinese digital library research output, 1994–2010. Journal of scientometrics, 89: 51– 87.
[26] Zhao, Yonghua, and Hong Lin. "WEB data mining applications in e-commerce.", IEEE 9th International Conference on Computer Science & Education, Vancouver, Canada, 2014.