



فرم پیشنهاد تحقیق پایان نامه‌ی کارشناسی ارشد

عنوان تحقیق به فارسی:

شخصی سازی نتایج موتورهای جستجو براساس تاریخچه رفتاری کاربر

نام:	دانشکده:
نام خانوادگی دانشجوی:	گروه تخصصی:
رشته تحصیلی:	گرایش:
سال ورود به مقطع جاری:	نیمسال ورودی:

نام و نام خانوادگی استاد (اساتید) راهنما:	نام و نام خانوادگی استاد (اساتید) مشاور:
-۱	-۱
-۲	-۲

این قسمت توسط حوزه معاونت پژوهش و فن آوری واحد تکمیل میگردد

تاریخ تأیید در شورای گروه: تاریخ دریافت حوزه پژوهشی واحد:

تاریخ تصویب در شورای تحصیلات تکمیلی ::

تأیید مدیر پژوهش و فن آوری

تأیید کارشناس پژوهشی

تأیید معاون پژوهش و فن آوری

توجه: در صورت خدشه و یا لاک گرفتگی در تاریخ های تصویب گروه و یا شورای تحصیلات تکمیلی فرم پیشنهاد تحقیق قابل پذیرش نیست .

به نام خدا

توجه: لطفاً این فرم با مساعدت و هدایت استاد راهنما تکمیل شود.

۱- اطلاعات مربوط به دانشجو:

نام: نام خانوادگی: شماره دانشجویی:
مقطع: رشته تحصیلی: گروه تخصصی:
گرایش: نام دانشکده: سال ورود به مقطع جاری:
نیمسال ورودی:
آدرس پستی در تهران:
تلفن ثابت محل سکونت: تلفن همراه: پست الکترونیک:
آدرس پستی در شهرستان:
تلفن ثابت محل سکونت: تلفن محل کار: دورنگار:

۲- اطلاعات مربوط به استاد راهنما:

تذکرات:

- دانشجویان دوره کارشناسی می توانند یک استاد راهنما و حداکثر دو استاد مشاور می توانند انتخاب نمایند.
- در صورتی که اساتید راهنما و مشاور **مدعو** می باشند، لازم است سوابق تحصیلی، آموزشی و پژوهشی کامل ایشان (رزومه کامل) شامل فهرست پایان نامه های کارشناسی ارشد و رساله های دکتری دفاع شده و یا در حال انجام که اساتید مدعو، راهنمایی و یا مشاوره آنرا بر عهده داشته اند، **به همراه مدارک مربوطه** و همچنین آخرین حکم کارگزینی (حکم هیأت علمی) ضمیمه گردد.
- اساتید راهنما و مشاور موظف هستند قبل از پذیرش پروپوزال، به سقف ظرفیت پذیرش خود توجه نموده و در صورت تکمیل بودن ظرفیت پذیرش، از ارسال آن به حوزه پژوهشی و یا در نوبت قرارداد و ایجاد وقفه در کار دانشجویان جداً پرهیز نمایند. بدیهی است در صورت عدم رعایت موازین مربوطه، مسئولیت تأخیر در ارائه پروپوزال و عواقب کار، به ترتیب متوجه اساتید راهنما، مشاور و گروه تخصصی خواهد بود.

اطلاعات مربوط به استاد راهنمای اول:

دانشگاهی

نام و نام خانوادگی: آخرین مدرک تحصیلی:
حوزوی

تخصص اصلی: رتبه دانشگاهی (مرتبه علمی): تلفن همراه:

تلفن منزل یا محل کار: پست الکترونیک (Email):

نحوه همکاری با واحد علوم و تحقیقات:

تمام وقت نیمه وقت مدعو

اطلاعات مربوط به استاد راهنمای دوم:

دانشگاهی

نام و نام خانوادگی: آخرین مدرک تحصیلی:
حوزوی

تخصص اصلی: رتبه دانشگاهی (مرتبه علمی): تلفن همراه:

تلفن منزل یا محل کار: پست الکترونیک (Email):

نحوه همکاری با واحد علوم و تحقیقات:

تمام وقت نیمه وقت مدعو

اطلاعات مربوط به استاد مشاور

دانشگاهی

نام و نام خانوادگی: آخرین مدرک تحصیلی:
حوزوی

تخصص اصلی: رتبه دانشگاهی (مرتبه علمی): تلفن همراه:

تلفن منزل یا محل کار: پست الکترونیک (Email):

نحوه همکاری با واحد علوم و تحقیقات:

تمام وقت نیمه وقت مدعو

اطلاعات مربوط به استاد مشاور دوم:

دانشگاهی

نام و نام خانوادگی: آخرین مدرک تحصیلی:
حوزوی

تخصص اصلی: رتبه دانشگاهی (مرتبه علمی): تلفن همراه:

تلفن منزل یا محل کار: پست الکترونیک (Email):

نحوه همکاری با واحد علوم و تحقیقات:

تمام وقت نیمه وقت مدعو

الف- عنوان تحقیق

۱- عنوان به زبان فارسی:

شخصی سازی نتایج موتورهای جستجو براساس تاریخچه رفتاری کاربر

۲- عنوان به زبان انگلیسی/(آلمانی، فرانسه، عربی):

تذکر: صرفاً دانشجویان رشته‌های زبان آلمانی، فرانسه و عربی مجازند عنوان پایان‌نامه خود را به زبان مربوطه در این بخش درج نمایند و برای بقیه دانشجویان، عنوان بایستی به زبان انگلیسی ذکر شود.

Personalizing Search Engine Result using user Intersets

ب- تعداد واحد پایان‌نامه:

ج- بیان مسأله اساسی تحقیق به طور کلی (شامل تشریح مسأله و معرفی آن، بیان جنبه‌های مجهول و مبهم، بیان متغیرهای مربوطه و منظور از تحقیق):

امروزه رشد و تغییر نمایی در حجم اطلاعات وب موجب عدم دسترسی سریع و آسان کاربران به اطلاعات مورد نیازشان، شده است. مشکل محققین امروز، کمبود منابع نیست بلکه برعکس، حجم عظیم اطلاعات، یک چالش بحث‌برانگیز شده است و پیدا کردن اطلاعات مرتبط و باکیفیت از بین انبوه موضوعاتی که در قالب‌های مختلف منتشر می‌شود نیز، چالشی دیگر محسوب می‌شود. اگرچه بکارگیری ابزارهایی مانند موتورهای جستجو می‌تواند به کاربران در یافتن اطلاعات مورد نیازشان کمک نماید و باعث تسریع و تسهیل دسترسی به اطلاعات وب شود، اما حجم اطلاعاتی که موتورهای جستجو به کاربر تحویل می‌دهند، معمولاً بسیار بیشتر از آن است که قابل پردازش سریع توسط کاربر باشد. علاوه بر این اغلب موتورهای جستجو از محتوای صفحات و ساختار پیوند بین آنها، برای تعیین میزان ارتباط صفحات با پرس وجوی کاربر استفاده می‌کنند. در واقع موتورهای جستجو، علایق و دانش پیش زمینه کاربر را در جستجو، مورد توجه قرار نمی‌دهند بلکه نتایج مشابهی را برای کاربران با علایق متفاوت و پرس وجوی یکسان بازایی می‌نمایند. بنابراین با توجه به اهمیت دسترسی سریع به اطلاعات مناسب برای کاربران در تعامل با وب، شخصی سازی کردن وب به یک پدیده محبوب به منظور سفارشی کردن محیط‌های وب تبدیل شده که نقش مهمی را در سرعت دسترسی و فراهم نمودن اطلاعات مورد نیاز کاربران بر طبق علایق آنها، ایفا می‌کند. در این میان شخصی سازی نتایج حاصل از موتورهای جستجوی وب، از اهمیت خاصی برخوردار می‌باشد. موتورهای جستجوی مبتنی بر کلمه کلیدی همچون Yahoo و Google ابزارهای اصلی برای استفاده از وب امروزی هستند. واضح است که وب بدون در نظر گرفتن موتورهای جستجو، نمی‌توانست به موفقیت عظیم کنونی دست یابد. اگرچه مشکلاتی جدی در رابطه با استفاده از آنها هم وجود دارد:

۱. فراخوانی بالا، دقت پایین؛ حتی اگر صفحات مرتبط اصلی هم، بازیابی شود؛ استفاده مناسبی از آن صفحات نمی‌توان نمود زیرا تعداد بازیابی زیاد اطلاعات، می‌تواند به بدی تعداد کم آنها باشد.
۲. فراخوانی کم یا صفر؛ اغلب اتفاق می‌افتد موتور جستجو، هیچ پاسخ مرتبطی برای پرس‌وجوی مورد نظر ارائه نمی‌دهد و یا اینکه صفحات مربوط و مهم بازیابی نمی‌شود.
۳. نتایج بسیار حساس به واژگانند؛ اغلب کلمات کلیدی ابتدایی ما به نتایج دلخواه منجر نمی‌شوند، در این صورت سندهای مرتبط دارای کلمات متفاوت با پرس‌وجو خواهد بود. این مساله باعث نارضایتی می‌شود زیرا پرس‌وجوهایی با معنای مشابه، باید پاسخ‌های همانند را بازگردانند.

جالب است که برخلاف پیشرفت‌هایی که در فناوری موتور جستجو صورت گرفته است، این مشکلات همچنان باقی مانده‌اند. در این تحقیق ما قصد داریم با ارائه یک روند منطقی، به کلاسه‌بندی رفتار کاربران و برچسب‌گذاری آنها به شیوه‌ای مناسب بپردازیم.

د - اهمیت و ضرورت انجام تحقیق (شامل اختلاف نظرها و خلاءهای تحقیقاتی موجود، میزان نیاز به موضوع، فواید احتمالی نظری و عملی آن و همچنین مواد، روش و یا فرآیند تحقیقی احتمالاً جدیدی که در این تحقیق مورد استفاده قرار می‌گیرد:

گرانبار شدن اطلاعات یک مشکل عمده در وب کنونی به شمار می‌رود. برای مقابله با این مشکل، سیستم‌های شخصی‌سازی وب ارائه شده‌اند که محتوا و سرویس‌های یک وب‌سایت را با افراد براساس علایق و رفتار گردشی آنها سازگار می‌کنند. از جمله دلایل نیاز به شخصی‌سازی وب، گرانبار شدن اطلاعات و نیاز به جذب مشتری پابرجا برای وب‌سایت‌ها می‌باشد.

وب جهانی، منبعی عظیم از اطلاعات را فراهم آورده است. در بررسی‌های گوناگون انجام شده در زمینه‌ی گسترش وب، تخمین زده شده است که روزانه بیش از یک میلیون صفحه به وب اضافه می‌شود و بیش از ۶۰۰ گیگابایت از صفحات در هر ماه تغییر می‌کنند [۱] و [۲]. این پدیده که گرانبار شدن اطلاعات نامیده می‌شود مشکلاتی را برای کاربران وب بوجود آورده است.

از مهم‌ترین این مشکلات عدم دسترسی آسان به اطلاعات مورد نیاز می‌باشد. در این انباره‌ی عظیم، کاربران در یافتن اطلاعات مورد نیاز خود در زمان مناسب و بصورت آسان دچار مشکل هستند؛ زیرا از یک سو باید میزان ربط هر صفحه را با نیاز خود بررسی کنند و از سوی دیگر باید صفحات را از نظر میزان قابلیت اعتماد ارزیابی کنند.

برای رفع این مشکل در دهه‌های اخیر سیستم‌های بازیابی اطلاعات و به تبع آن موتورهای جستجو ایجاد شدند که محتوای صفحات وب را اندیس‌گذاری می‌کنند و صفحات مرتبط با پرسش کاربر را برمی‌گردانند، اما این تکنولوژی‌ها دو مشکل دارند:

- حجم اطلاعاتی که موتورهای جستجو به کاربر تحویل می‌دهند بسیار بیشتر از آن است که قابل پردازش توسط کاربر باشد.
- بیشتر موتورهای جستجوی معروف موجود در وب، تنها از محتوای صفحات و ساختار ارتباطی موجود بین آن‌ها برای ارزیابی میزان ربط صفحات با پرسش کاربر استفاده می‌کنند. از این رو برایشان اهمیت ندارد که پرسش توسط چه کسی و با چه علایق و دانش پیش‌زمینه‌ای وارد شده است. نتیجه آنکه صفحات بازاریابی شده برای دو کاربر با علایق متفاوت و با پرسش یکسان مجموعه‌ی یکسانی از صفحات وب می‌باشد.

هدف یک سیستم شخصی‌سازی وب، فراهم کردن اطلاعات دلخواه یا مورد نیاز کاربران بدون درخواست صریح آن‌ها می‌باشد. از فواید سیستم شخصی‌سازی وب می‌توان به موارد زیر اشاره نمود:

۱. نقش مهمی در کاهش گرانباز شدن اطلاعات ایفا می‌کند و وب‌سایت را به یک محیط کاربرپسندتر برای افراد تبدیل می‌کند.
۲. با فراهم کردن اطلاعات دلخواه کاربر به روش مناسب و در زمان مناسب باعث بهبود گردش کاربر در وب‌سایت می‌شود.
۳. در تجارت الکترونیکی، مکانیسمی برای درک بهتر نیازهای مشتری، شناسایی تمایلات آینده‌ی او و در نهایت افزایش پابرجایی مشتری به سرویس ارائه شده، فراهم می‌کند.

رویکرد مبتنی بر کاربرد در شخصی‌سازی وب یک عیب مهم دارد و آن این است که فرآیند توصیه به کاربر تنها براساس داده‌های تراکنشی موجود او صورت می‌گیرد و از این رو اقلام یا صفحاتی که اخیراً به سایت اضافه شده اند نمی‌توانند به او توصیه شوند. این مشکل عموماً مشکل قلم جدید نامیده می‌شود. از سوی دیگر اگرچه الگوهای کشف شده‌ی مربوط به وب در کشف ارتباطات اقلام با یکدیگر یا کاربران با یکدیگر مفیدند، اما بدون استفاده از دانش عمیق تری از دامنه‌ی وب‌سایت‌ها، درک اندکی از دلایل آن که چرا اقلام یا کاربران در گروه‌هایی با هم قرار می‌گیرند، در اختیار ما قرار می‌دهند. یک رویکرد معمول برای حل این مشکل در فیلتر کردن جمعی آن است که مشخصات محتوای صفحات را با رتبه‌بندی‌ها و قضاوت‌های کاربر ادغام کنیم. بطور کلی در این رویکردها کلمات کلیدی از محتوای وب‌سایت استخراج می‌شوند و برای اندیس‌گذاری صفحات براساس محتوا یا طبقه‌بندی آن‌ها به دسته‌های مختلف، مورد استفاده قرار می‌گیرند. در حوزه‌ی شخصی‌سازی وب این رویکرد به سیستم اجازه می‌دهد تا صفحات را نه تنها براساس افراد مشابه، بلکه براساس شباهت محتوایی آن‌ها به صفحاتی که کاربر اخیراً بازدید کرده است به او توصیه کند.

شخصی‌سازی موتور جستجو یکی از انواع شخصی‌سازی‌های عمقی محسوب می‌شود. می‌توان گفت سیستم‌های شخصی‌سازی که به کاربران در یافتن اطلاعات مورد نیازشان کمک می‌نمایند نیازمند استفاده از تکنیک‌های تحلیل اطلاعات متنی و معنایی می‌باشند که در زمینه بازاریابی اطلاعات وجود دارند. از اواخر سال ۱۹۹۰ عبارت "شخصی‌سازی" عمومیت یافت [۳]. شخصی‌سازی وب با هدف فراهم کردن اطلاعات متناسب

با علایق کاربران نقش مهمی را در بهبود تعامل کاربران با وب داشته است. از آنجاییکه امروزه موتورهای جستجو ابزار اصلی جستجوی اطلاعات موجود بر روی وب می‌باشند و علی‌رغم استفاده گسترده نمی‌توانند نتایج جستجوی مناسب با نیاز بسیاری از کاربران را فراهم نمایند [۴]، بنابراین شخصی سازی موتورهای جستجو با هدف کمک به کاربران در یافتن اطلاعات، متناسب با علایق آنها ضروری به نظر می‌رسد.

در زمینه شخصی سازی جستجو تاکنون کارهای متعددی صورت گرفته و در هر یک از کارهای انجام شده رویکرد محققان نسبت به این مسئله و آنچه بدان پرداخته‌اند متفاوت بوده است. این شخصی سازی بر اساس پروفایلی از کاربر انجام می‌شود، اینکه این پروفایل شامل چه داده‌هایی بوده و چگونه مقاردهی اولیه، بروزرسانی و استفاده می‌گردد منجر به ایجاد روش‌های متعددی در این زمینه گشته است. شخصی سازی وب، فرآیند بهینه نمودن روش‌های دستیابی به صفحات وب است، بطوریکه منطبق با نیاز هر کاربر می‌باشد. آنچه که راه‌گشای مشکل فعلی موتورهای جستجو و باعث تسریع در عملکرد آنها می‌باشد، ارائه چارچوبی مناسب جهت کشف الگوی صحیح با توجه به اقلام بزرگ موجود در تاریخچه‌ی کاربران می‌باشد. این رویکرد باعث بهبود در روند توصیه کننده‌های موتورهای جستجو نیز می‌شود.

۵- مرور ادبیات و سوابق مربوطه (بیان مختصر پیشینه تحقیقات انجام شده در داخل و خارج کشور پیرامون موضوع تحقیق و نتایج آنها و مرور ادبیات و چارچوب نظری تحقیق):

در سال‌های اخیر مسائل و چالش‌های جدیدی در زمینه‌ی شخصی سازی موتورهای جستجو بر مبنای داده کاوی مطرح شده است که از جمله‌ی مهم ترین آنها می‌توان به کاوش پروفایل‌های تکاملی و نیز وارد کردن دانش دامنه در فرآیند شخصی سازی وب اشاره کرد.

در [۳] به بررسی سیستم‌های تعاملی متعددی پرداخته شده و به این نتیجه رسیده است که یادگیری الگوهای رفتاری تک‌تک کاربران و انطباق سیستم با این الگوهای رفتاری، کار بسیار ارزشمندی است. یک مثال آشنا سایت *amazon.com* می‌باشد که براساس دفعات قبلی که کاربران از سایت بازدید کرده‌اند پیشنهاداتی را پیش روی آنها قرار می‌دهد.

در [۵] یک روشی برای استخراج علاقه‌مندی کاربران ارائه شده است. در این مقاله مجموعه‌ی علاقه‌مندی‌های هر کاربر به صورت صریح ذخیره شده و شش نوع پروفایل مختلف برای کاربران با یک تابع فاصله تعریف شده که در نهایت سه توصیه با چهار نرخ مختلف به هر کاربر واگذار می‌شود.

در [۶] ابتدا تراکنش‌های مورد علاقه‌ی کاربران با استفاده از الگوریتم استخراج ویژگی روی لاگ فایل کاربران، ذخیره می‌شوند سپس آنها با استفاده از یک روش خوشه‌بندی جدید با در نظر گرفتن یک حد آستانه برای الویت دادن به اقلام بزرگ، تراکنش‌ها خوشه‌بندی می‌شوند. این خوشه‌ها در بهبود رتبه بندی نتایج موثر و همچنین جستجوی دقیق‌تری ارائه می‌کند.

در [۷] یک موتور جستجوی مکان آگاه بر اساس الگوریتم‌های یادگیری ماشین به منظور افزایش دقت جستجو ارائه شده است. این الگوریتم با در نظر گرفتن آدرس IP نتایج جستجو را اعلام می‌کند. نتایج تحقیقات این پژوهش نشان داد این روش سریعترین موتور جستجوی شخصی محسوب می‌شود.

در [۸] ابتدا یک پایگاه داده ضمنی از تراکنش کاربران تولید می‌شود؛ سپس با توجه به مفاهیم سلسله مراتبی، داده‌ها طبقه بندی می‌شوند در نهایت الگوریتم رتبه‌بندی مجدداً اجرا می‌شود. این روش باعث بهبود ۳۴٪ ای در دقت نتایج می‌شود.

در [۹] اثر یادگیری را بر روی الگوهای رفتاری اشخاص نمایش می‌دهد و یک رویکرد شخصی سازی موتور جستجو با در نظر گرفتن تاریخچه رفتاری کاربر بدون تغییر آن برای هر کاربر، بیان می‌شود؛ چرا که الگوی رفتاری کاربر یک روال معینی دارد که قابل یادگیری توسط ماشین است.

در [۱۱]، یک رویکرد مبتنی بر عامل برای مسئله شخصی سازی جستجو با استفاده از قابلیت‌های هوشمندی، یادگیری و هدف‌گرایی، پیشنهاد شده است. هدف از این پژوهش، ارائه چارچوبی جهت مقایسه سیستم‌های شخصی سازی طراحی شده و بررسی سیستم‌های طراحی شده برای نابینایان می‌باشد.

در [۱۲]، روش جدیدی برای شخصی سازی نتایج موتورهای جستجو با استفاده از شبکه‌های مفهومی فازی سازگار اتوماتیک ارائه شده است. در این روش با در نظر گرفتن پروفایل تعدادی از کاربران مورد نظر، یک الگوریتم خوشه بندی بر روی آنها اعمال می‌شود، سپس یک شبکه مفهومی فازی برای هر خوشه بصورت اتوماتیک ساخته می‌شود. نتایج جستجوی کاربر جدید نیز با یک شبکه مفهومی فازی نمایش داده می‌شود.

و - جنبه جدید بودن و نوآوری در تحقیق:

با شناسایی الگوی رفتاری کاربر بر مبنای استفاده از خوشه‌بندی و برچسب‌گذاری مناسب بر روی اطلاعات تراکنشی موجود در پایگاه داده، به منظور تحلیل معنایی صفحاتی که در تاریخچه رفتاری کاربر ثبت شده است، سعی در بهبود حل مسئله داریم.

ز- اهداف مشخص تحقیق (شامل اهداف آرمانی، کلی، اهداف ویژه و کاربردی):

اهداف کلی:

در حال حاضر بیش از سه میلیارد صفحه وب در برابر کاربران وب قرار دارد و هر وب‌سایت هم سعی در جذب کاربران بیشتر و افزایش طول مدت جلسات کاربران، دارد. برای جذب کاربران بیشتر نیاز به سرویس‌های جدیدی است که نیاز کاربران را تشخیص داده و بتواند کیفیت سرویس^۱ بهتری را ارائه نماید تا باعث جذب بیشتر مخاطب شود. هدف اصلی از این پژوهش، بهبود مدل رفتاری کاربران بر اساس تاریخچه پرس و جوهای آنها، با کیفیت بیشتر نسبت به پرس و جوهای که بدون در نظر گرفتن دانش آنها تولید می‌شود، می‌باشد. لذا مطلوب است که تا حد امکان این فرآیند بصورت خودکار انجام شود.

¹ Quality of Service

اهداف ویژه:

۱. کاوش در داده های رفتاری کاربران جهت آشنا شدن با علایق آنها
۲. بهبود وضعیت ساختاری وبسایت جهت جذب کاربران
۳. ارائه نمودن لینک های مورد علاقه کاربر
۴. شخصی سازی اطلاعات بعد از درک علاقمندی های کاربر
۵. شخصی سازی نتایج جستجو بعد از درک علاقه مندی های کاربر

ح - در صورت داشتن هدف کاربردی، نام بهره‌وران (سازمان‌ها، صنایع و یا گروه ذینفعان) ذکر شود (به عبارت دیگر محل اجرای مطالعه موردی):

۱. کلیه وبسایت‌ها
۲. موسسات آموزشی و تحقیقاتی
۳. کلیه انجمن‌ها و کانون‌ها
۴. مراکز خرید و فروش اجناس
۵. کلیه سازمان‌ها ادارات
۶. کتابخانه‌ها دیجیتال

ط - سؤالات تحقیق:

۱. برای تعامل با کاربران جهت استخراج الگوهای رفتاری آنها چه تکنولوژی‌های جدیدی وجود دارد؟
۲. از چه الگوریتم‌هایی می‌توان جهت بهبود وضعیت ساختاری وبسایت‌ها با در نظر گرفتن وضعیت رفتاری کاربران، کاوش نمود؟
۳. آیا با استفاده از الگوریتم‌های وب کاوی می‌توان به علائق کاربران پی برد؟

۱. تشخیص الگویی مناسب جهت شناسایی علاقمندی‌های کاربران، نقش مهمی در جذب کاربران دارد.

۲. با درک علاقمندی کاربران، می‌توان نتایج جستجو را با در نظر گرفتن آن تغییر داد و نتیجه کار را بهبود بخشید.

۳. با توجه به علاقمندی کاربران، می‌توان پیشنهادهای معقول‌تری به آنها و در نتیجه ائتلاف زمان کمتری برای آنها به ارمغان آورد.

ک- تعریف واژه‌ها و اصطلاحات فنی و تخصصی (به صورت مفهومی و عملیاتی):

شخصی‌سازی وب: هر اقدامی که اطلاعات یا سرویس‌های فراهم شده توسط یک وب‌سایت را با نیازهای یک کاربر یا گروه خاصی از کاربران با بکارگیری دانش بدست آمده از رفتارگردشی کاربر و علائق خاص او بصورت ترکیب با محتوا و ساختار وب‌سایت سازگار کند، شخصی‌سازی وب نامیده می‌شود [۱۰].

وب کاوی: وب کاوی، به کارگیری تکنیک‌های داده کاوی^۲ برای کشف و استخراج خودکار اطلاعات از اسناد و سرویس‌های وب می‌باشد.

موتور جستجو: خود یک وب‌سایت اینترنتی است که به‌عنوان یکی از ابزارهای بازیابی اطلاعات رایگان مطرح است. وظیفه آنها جمع‌آوری اطلاعات از سطح اینترنت و راهنمایی کاربران به سوی سایت‌های مورد نظر است.

خوشه بندی: خوشه بندی یکی از شاخه‌های یادگیری بدون نظارت می‌باشد. فرآیند خودکاری است که در طی آن، نمونه‌ها به دسته‌هایی که اعضای آن مشابه یکدیگر می‌باشند تقسیم می‌شوند، که به این دسته‌ها خوشه^۳ گفته می‌شود. بنابراین خوشه مجموعه‌ای از اشیاء می‌باشد که در آن اشیاء با یکدیگر مشابه بوده و با اشیاء موجود در خوشه‌های دیگر غیر مشابه می‌باشند.

قوانین انجمنی^۴: قوانین انجمنی، روابط و وابستگی‌های متقابل بین مجموعه بزرگی از اقلام داده‌ای را نشان می‌دهند.

² Data Mining

³ Cluster

⁴ Association Rules

الف- شرح کامل روش تحقیق بر حسب هدف، نوع داده ها و نحوه اجراء (شامل مواد، تجهیزات و استانداردهای مورد استفاده در قالب مراحل اجرایی تحقیق به تفکیک):

تذکر: درخصوص تفکیک مراحل اجرایی تحقیق و توضیح آن، از به کار بردن عناوین کلی نظیر، «گردآوری اطلاعات اولیه»، «تهیه نمونه‌های آزمون»، «انجام آزمایش‌ها» و غیره خودداری شده و لازم است در هر مورد توضیحات کامل در رابطه با منابع و مراکز تهیه داده‌ها و ملزومات، نوع فعالیت، مواد، روش‌ها، استانداردها، تجهیزات و مشخصات هر یک ارائه گردد.

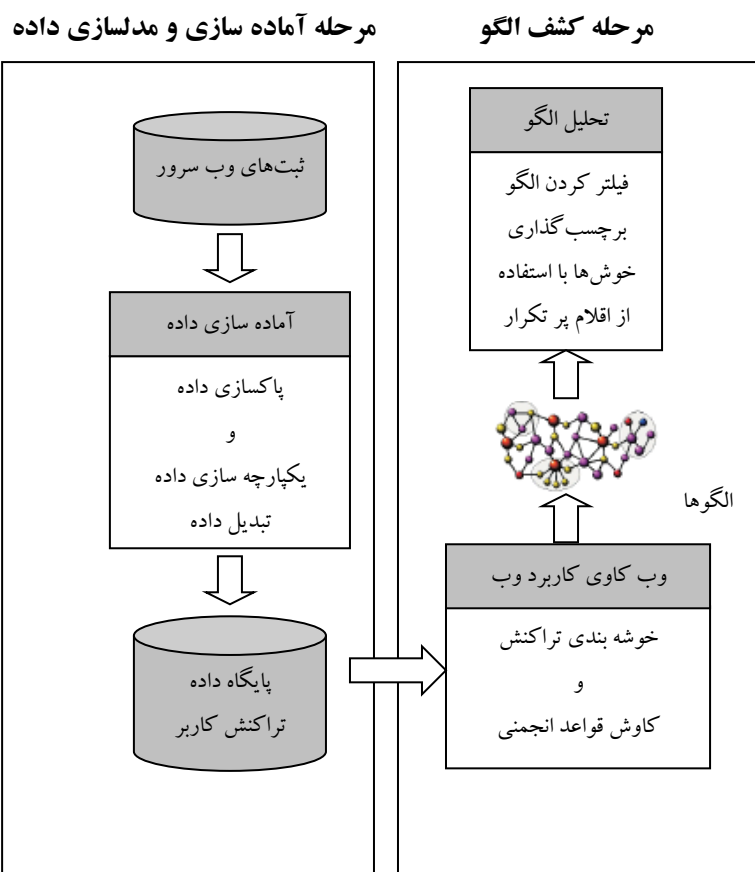
منابع عمده‌ی داده در موتورهای جستجو، وب فایل‌های ثبت وب سرور می‌باشند که شامل ثبت‌های دسترسی وب سرور و ثبت‌های سرویس‌دهنده‌ی خدمات وب است. در ثبت گرفتن وب سرور، داده‌ها توسط وب سرور به یک یا چند فرمت ثبت می‌شود.

داده‌های وب جمع‌آوری شده معمولاً دارای حجم زیاد و بسیار ناهمگن می‌باشند. این داده‌ها باید به داده‌های سازگار و یکپارچه تبدیل شوند تا بتوانند برای مرحله‌ی کشف الگو مفید باشند.

مرحله‌ی اول، مرحله آماده سازی داده می‌باشد. ثبت‌های خام وب را به داده‌ی تراکنشی تبدیل می‌کند که می‌تواند در داده کاوی مورد استفاده قرار گیرد. این مرحله همچنین شامل یکپارچه سازی داده از منابع مختلف مانند پایگاه‌های داده، سرویس‌های خدمات کاربردی و محتوای سایت می‌باشد.

مرحله‌ی دوم، مرحله خوشه‌بندی داده‌ها می‌باشد. روش‌های مختلف داده کاوی مورد نیاز در این تحقیق در این بخش اجرا می‌شود. ابتدا خوشه‌بندی روی داده‌های تراکنشی اعمال می‌شود. هر خوشه با توجه به مجموعه اقلام بزرگ موجود در تاریخچه رفتاری کاربر، اندیس‌گذاری می‌شود. در این مرحله، داده‌ها به کلاس‌هایی تقسیم می‌شوند که بیشترین شباهت را به هم دارند لذا گروه‌هایی شبیه به هم با برجسب شباهت‌های آشکار(اقلام بزرگ)، ساخته می‌شوند.

مرحله‌ی سوم، مرحله کشف الگوهای مهم با در نظر گرفتن یک حد آستانه می‌باشد. نتایج این مرحله به پروفایل‌های تجمعی کاربر تبدیل می‌شوند که برای استفاده در بخش توصیه مناسب می‌باشد.



ب- متغیرهای مورد بررسی در قالب یک مدل مفهومی و شرح چگونگی بررسی و اندازه گیری متغیرها:

متغیرهایی مانند تعداد کاربران، صفحات مورد بررسی، وبسایت‌های مورد مراجعه و علائق کاربران مورد بررسی قرار گرفته‌اند و عملیات وب کاوی امکان شمارش پارامترها را فراهم می‌کند.

ج - شرح کامل روش (میدانی، کتابخانه‌ای) و ابزار (مشاهده و آزمون، پرسشنامه، مصاحبه، فیش‌برداری و غیره) گردآوری داده‌ها:

ابتدا با مطالعه منابع کتابخانه‌ای به بررسی روند کلی کار، کارهای انجام شده، بررسی چارچوب‌ها و الگوریتم‌های موجود در زمینه موتورهای جستجو می‌پردازیم.

فرآیند کلی شخصی سازی وب براساس وب کاوی شامل سه مرحله است:

۱. آماده سازی داده‌ها (پیش پردازش داده)

۲. کشف الگوی موجود در بین داده های آماده شده از مرحله قبل (مدلسازی)

۳. تجزیه و تحلیل الگوهای بدست آمده و استفاده از الگوهای کشف شده برای شخصی سازی وب.

در فاز اول، پردازش ابتدایی داده‌ها صورت می‌گیرد. در فاز دوم از متدهای آماری، از جمله متدهای ویژه داده کاوی به منظور کشف الگوی مورد نظر، استفاده می‌شود.

د - جامعه آماری، روش نمونه‌گیری و حجم نمونه (در صورت وجود و امکان):

تکنیک‌های مهمی وجود دارد که هنگام انجام عمل پیش‌پردازش باید مورد توجه قرار گیرند. آماده‌سازی داده‌های ثبت وب سرورها برای اینکه بتواند در ادامه فرایند تشخیص الگو استفاده شوند، بسیار ضروری به نظر می‌رسد. ورودی‌های مرحله پیش‌پردازش داده‌های خام فایل‌های ثبتی بوده و خروجی این فایل‌ها User-Session نامیده می‌شود.

این مرحله خود شامل گام‌های زیر است که هر یک وظیفه خاصی را جهت پردازش داده‌ها بر عهده دارند.

۱. پاکسازی داده‌ها

۲. تشخیص کاربر

۳. تشخیص Session

۴. تشخیص صفحه مشاهده شده

۵. تکمیل مسیر

در عملیات وب‌کاوی این مراحل بصورت استاندارد بوده و نحوه انجام آن شفاف می‌باشد.

این اطلاعات خروجی در نهایت به صورت تراکنش ذخیره می‌شوند. برای انجام آزمون‌ها و مشاهدات، از نرم‌افزار MATLAB استفاده می‌کنیم.

ه - روش‌ها و ابزار تجزیه و تحلیل داده‌ها:

ابتدا با استفاده از روش‌های خوشه‌بندی داده‌های تراکنشی، علایق کاربران را به دسته‌هایی مختلف تقسیم می‌نماییم. با اعمال این الگوریتم‌ها کاربران بر اساس نوع استفاده از اینترنت به گروه‌های مختلفی تقسیم می‌شوند. سپس با استفاده از الگوریتم‌های انجمنی الگوی استفاده از سایت‌های مختلف و کلمات جستجو شده برای هر خوشه از کاربران که دارای علایق مشابهی هستند را بدست می‌آوریم سپس با توجه به آمار اقلام مورد علاقه کاربران، خوشه‌ها اندیس گذاری می‌شوند. لذا در این تحقیق از ابزارهای خوشه‌بندی و کشف قوانین انجمنی استفاده می‌شود. بکارگیری ابزارها و الگوریتم‌های جدید مانند الگوریتم‌های یادگیری ماشین و ابزارهای داده‌کاوی در شخصی سازی موتورهای جستجو می‌تواند باعث کارایی عملکرد آنها شود.

1. P. Achananuparp, H. Han, O. Nasraoui and R. Johnson, "Semantically Enhanced User Modeling", Proceedings of the 2007 ACM Symposium on Applied Computing (Seoul, Korea, March 11 - 15, 2007). SAC '07. ACM, New York, NY, 1335-1339.
2. O. Nasraoui, M. Soliman, E. Saka, A. Badia and R. Germain, "A Web Usage Mining Framework for Mining Evolving User Profiles in Dynamic Web Sites", IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, 2008, vol. 20, 1041-4347.
3. A. Jameson, "Adaptive interfaces and agents," Human-Computer Interaction: Design Issues, Solutions, and Applications, 2009, p. 105-131.
4. W. Alhalabi, M. Kubat, M. Tapia, "Search Engine Personalization Tool Using Linear Vector Algorithm", Proceedings of the 4th Saudi Technical Conference and Exhibition 2006 (336-344).
5. Y. Wu, Y. Chen, L. P. Chen, "Enabling Personalized Recommendation on the Web Based on User Interests and Behaviors", Research Issues in Data Engineering, Proceedings 2001 (17 - 24).
6. X. Chen, L. Huang, "The Research of Personalized Search Engine Based on Users' access Interest", Computational Intelligence and Industrial Applications, 2009(337 - 340).
7. M. Sun, W. Su, L. Shu, M. Li, L. Xue, "FoSSicker: A personalized search engine by location-awareness", Computing, Networking and Communications (ICNC), 2012(456 - 460).
8. M. Speretta, "Personalizing Search Based on User Search Histories", Thesis for the degree of Master of Science, Defense Date: June 17, 2005.
9. D. Savenkov, D. Lagun, Q. Liu, "Search Engine Switching Detection Based on User Personal Preferences and Behavior Patterns", Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval, 2013 (33-42)
10. M. Eirinaki and M. Vazirgiannis, "Web Mining for Web Personalization", ACM Transactions on Internet Technology, 3:1-27, 2003.

۱۱. الهام عبدنیکویی پور، "تعامل انسان-کامپیوتر مبتنی بر عامل برای نابینایان"، پایان نامه کارشناسی ارشد

مهندسی کامپیوتر گرایش نرم افزار دانشگاه صنعتی امیرکبیر، اسفند ۸۹

۱۲. بتول ارزانیان، "شخصی سازی موتورهای جستجو با استفاده از روشهای هوشمند"، پایان نامه کارشناسی

ارشد مهندسی کامپیوتر گرایش هوش مصنوعی دانشگاه کردستان، دی ۸۸

۶- استفاده از امکانات آزمایشگاهی واحد:

آیا برای انجام تحقیقات نیاز به استفاده از امکانات آزمایشگاهی واحد علوم و تحقیقات می‌باشد؟ بلی خیر
 در صورت نیاز به امکانات آزمایشگاهی لازم است نوع آزمایشگاه، تجهیزات، مواد و وسایل مورد نیاز در این قسمت مشخص گردد.

نوع آزمایشگاه	تجهیزات مورد نیاز	مواد و وسایل	مقدار مورد نیاز

امضاء استاد راهنما:

امضاء مدیر گروه تخصصی:

۷- زمان بندی انجام تحقیق:

الف- تاریخ شروع:..... ب- مدت زمان انجام تحقیق:..... ج- تاریخ اتمام:.....

تذکر: لازم است کلیه فعالیت‌ها و مراحل اجرایی تحقیق (شامل زمان ارائه گزارشات دوره‌ای) و مدت زمان مورد نیاز برای هر یک، به تفکیک پیش‌بینی و در جدول مربوطه درج گردیده و در هنگام انجام عملی تحقیق، حتی‌الامکان رعایت گردد.

پیش‌بینی زمان‌بندی فعالیت‌ها و مراحل اجرایی تحقیق و ارائه گزارش پیشرفت کار

۱۵-۷/۱۰

زمان اجرا به ماه												زمان کل (ماه)	شرح فعالیت				
۱۲	۱۱	۱۰	۹	۸	۷	۶	۵	۴	۳	۲	۱						
																	۱
																	۲
																	۳
																	۴
																	۵
																	۶
																	۷
																	۸
																	۹
																	۱۰
																	۱۱
																	۱۲

توجه: ۱- زمان و نوع فعالیت‌های اجرایی پایان‌نامه، حتی‌الامکان باید با مندرجات جدول منطبق باشد.

۲- حداقل زمان قابل قبول برای پیش‌بینی مراحل مطالعاتی و اجرایی پایان‌نامه کارشناسی ارشد ۶ ماه و حداکثر ۱۲ ماه می‌باشد.

۸- صورت جلسه گروه تخصصی

نام و نام خانوادگی دانشجو: امضاء تاریخ

نام و نام خانوادگی استاد یا استادان راهنما امضاء تاریخ

-۱

-۲

نام و نام خانوادگی استاد یا استادان مشاور امضاء تاریخ

-۱

-۲

نام و نام خانوادگی عضو کمیته نظارت بر تحقیق امضاء تاریخ

-۱

شورای گروه تخصصی در تاریخ در محل با حضور اعضای مربوطه

خانم

تشکیل و موضوع پایان نامه با عنوان

آقای

بررسی و به تصویب رسید.

نام و نام خانوادگی اعضای شورا امضاء تاریخ

-۱

-۲

-۳

-۴

-۵

-۶

نام و نام خانوادگی مدیر گروه: امضاء تاریخ

توجه: در صورت خدشه و یا لاک گرفتگی در تاریخ های تصویب گروه و یا شورای تحصیلات تکمیلی فرم پیشنهاد تحقیق قابل پذیرش نیست.

۹- صورتجلسه شورای تحصیلات تکمیلی:

موضوع و طرح تحقیق پایان نامه آقای / خانم دانشجوی مقطع کارشناسی ارشد گروه
 که در تاریخ به تصویب کمیته گروه تخصصی مربوطه رسیده است:

۱- در جلسه مورخ شورای تحصیلات تکمیلی مطرح و:

- الف- به تصویب رسید.
- ب- مقرر گردید به دلیل اشکالات مطرح شده به گروه اعاده شود.
- ج- به تصویب نرسید.

۲- پروپوزال اصلاح شده در جلسه مورخ شورای تحصیلات تکمیلی مطرح و:

- الف- به تصویب رسید.
- ب- مقرر گردید با توجه به اینکه اشکالات مطرح شده رفع نشده است مجدداً به گروه اعاده شد.

۳- پروپوزال پس از رفع اشکالات مجدداً در جلسه مورخ شورای تحصیلات تکمیلی مطرح و به تصویب رسید.

ردیف	نام و نام خانوادگی	نوع رأی (موافق یا مخالف)	محل امضاء	توضیحات
۱				
۲				
۳				
۴				
۵				

نام و نام خانوادگی کارشناس پژوهشی: امضاء تاریخ

نام و نام خانوادگی مدیر پژوهش و فن آوری: امضاء تاریخ

نام و نام خانوادگی معاون پژوهش و فن آوری دانشگاه: امضاء تاریخ

این فرم باید توسط دانشجو تکمیل شود

فرم سازمان مرکزی دانشگاه آزاد اسلامی

فرم الف - فرم اطلاعات پایان نامه کارشناسی ارشد

نام واحد دانشگاهی: واحد علوم و تحقیقات خوزستان	
عنوان پایان نامه کارشناسی ارشد:	
نام و نام خانوادگی دانشجو:	نیمسال تحصیلی:
شماره دانشجویی:	تعداد واحد پایان نامه:
رشته تحصیلی:	گرایش:
<input type="checkbox"/> فنی و مهندسی	<input type="checkbox"/> علوم انسانی
<input type="checkbox"/> کشاورزی	<input type="checkbox"/> هنر
<input type="checkbox"/> کد رشته:	<input type="checkbox"/> علوم پایه
نام و نام خانوادگی استاد راهنما ۱:	رشته تحصیلی:
مرتبه علمی: استادیار <input type="checkbox"/> دانشیار <input type="checkbox"/> استاد <input type="checkbox"/>	کد شناسایی استاد راهنما:
نام و نام خانوادگی استاد راهنما ۲:	رشته تحصیلی:
مرتبه علمی: استادیار <input type="checkbox"/> دانشیار <input type="checkbox"/> استاد <input type="checkbox"/>	کد شناسایی استاد راهنما:
نام و نام خانوادگی استاد مشاور ۱:	رشته تحصیلی:
مرتبه علمی: استادیار <input type="checkbox"/> دانشیار <input type="checkbox"/> استاد <input type="checkbox"/> مربی <input type="checkbox"/>	کد شناسایی استاد راهنما:
نام و نام خانوادگی استاد مشاور ۲:	رشته تحصیلی:
مرتبه علمی: استادیار <input type="checkbox"/> دانشیار <input type="checkbox"/> استاد <input type="checkbox"/> مربی <input type="checkbox"/>	کد شناسایی استاد راهنما:

امضاء مدیر پژوهش دانشکده

نام و نام خانوادگی دانشجو :	رشته / گرایش :	سال ورود :
مقطع تحصیلی :	کارشناسی ارشد	
عنوان پروپوزال :		
استاد راهنمای اول :	استاد راهنمای دوم :	
استاد مشاور اول :	استاد مشاور دوم :	
<u>نظرات گروه تخصصی :</u>		
<u>نتیجه نهایی بررسی پروپوزال</u>		
<input type="checkbox"/> پروپوزال مورد تصویب قرار گرفت .		
<input type="checkbox"/> مقرر شد پس از انجام تصحیحات لازم , مجدداً" در گروه مطرح شود .		
<input type="checkbox"/> مقرر شد پس از انجام تصحیحات لازم و تایید مدیر گروه مورد تصویب قرار گیرد .		
<input type="checkbox"/> مورد تصویب قرار نگرفت .		
امضاء حاضرین در جلسه:		
۱-	۴-	
۲-	۵-	
۳-	۶-	
امضاء مدیر گروه تخصصی :		
امضاء نماینده پژوهش :		

نام و نام خانوادگی دانشجو :
رشته / گرایش :

ادامه نظرات گروه تخصصی :