

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ



دانشگاه کردستان
دانشکده فنی و مهندسی
گروه مهندسی کامپیوتر و فناوری اطلاعات

عنوان:

شخصی سازی موتورهای جستجو با استفاده از روش های هوشمند

پژوهشگر:

بتول ارزانیان

استاد راهنما:

دکتر فردین اخلاقیان

استاد مشاور:

دکتر پرهام مرادی

پایان نامه کارشناسی ارشد رشته مهندسی کامپیوتر گرایش هوش مصنوعی

دی ماه 1388

کلیه حقوق مادی و معنوی مترتب بر نتایج مطالعات،

ابتکارات و نوآوری های ناشی از تحقیق موضوع

این پایان نامه (رساله) متعلق به دانشگاه کردستان است.

***** تعهد نامه *****

اینجانب بتول ارزانیان دانشجوی کارشناسی ارشد رشته مهندسی کامپیوتر گرایش هوش مصنوعی دانشگاه کردستان، دانشکده فنی و مهندسی گروه مهندسی کامپیوتر و فناوری اطلاعات تعهد می نمایم که محتوای این پایان نامه نتیجه تلاش و تحقیقات خود بوده و از جایی کپی برداری نشده و به پایان رسانیدن آن نتیجه تلاش و مطالعات مستمر اینجانب و راهنمایی و مشاوره اساتید بوده است.

با تقدیم احترام

بتول ارزانیان

1388 /10/28



دانشگاه کردستان
دانشکده فنی و مهندسی
گروه مهندسی کامپیوتر و فناوری اطلاعات

پایان نامه کارشناسی ارشد رشته مهندسی کامپیوتر گرایش هوش مصنوعی

عنوان:

شخصی سازی موتورهای جستجو با استفاده از روش های هوشمند

پژوهشگر:

بتول ارزانیان

در تاریخ 1388/10/19 توسط کمیته تخصصی و هیات داوران زیر مورد بررسی قرار گرفت و با نمره.....
و درجه به تصویب رسید.

امضاء	مرتبه علمی	نام و نام خانوادگی	هیات داوران
	استادیار	دکتر فردین اخلاقیان	1-استاد راهنما
	استادیار	دکتر پرهام مرادی	2-استاد مشاور
	استادیار	دکتر محمد صنیعی	3-استاد داور خارجی
	استادیار	دکتر فردین احمدی زر	4-استاد داور داخلی

مهر و امضاء معاون پژوهشی و تحصیلات تکمیلی دانشکده

مهر و امضاء گروه

"تقدیم"

تقدیم به پدر و مادرم که به من چگونه زیستن را آموختند

تقدیم به آنان که دعای خیرشان بدرقه راهم بود

تقدیم به آنان که مشوق راه دانشم بودند

تقدیم به آنان که در راه کسب علم و معرفت برای من آنچه در توان داشتند انجام دادند

تقدیم به آنان که دوری مرا تحمل کردند و تقدیم به آنان که در رهگذر عمر یاری گر و دلگرمی من بودند

امیدوارم بتوانم ادای دین کنم و به خواسته آنان جامه عمل بپوشانم

خدایا عاقبت به خیری و عافیت و طول عمر را برای آنان از درگاہت مسئلت دارم.

"سپاسگزاری و قدردانی"

شکر خداوند متعال را به جای آورده که توفیق نصیب من کرد تا این پایان نامه را به پایان برسانم. در آغاز لازم می دانم از زحمات پدر و مادر عزیزم که همواره در طول زندگیم یار و یاورم بودند و دلگرمی های امیدبخش خواهران و برادر مهربانم کمال تشکر را بنمایم.

از استاد راهنمایم جناب آقای دکتر فریدین اخلاقیان که با راهنمایی های خود راهگشای اینجانب بوده اند و استاد مشاورم جناب آقای دکتر پرهام مرادی کمال تشکر و سپاسگزاری را دارم. از کلیه دوستانم که به هنگام نیاز از هیچگونه همدردی و مساعدتی دریغ نمودند، تقدیر و تشکر می نمایم.

در پایان نیز از جناب آقای مهندس آرمین سجادی، جناب آقای مهندس آسو حمزه ای، سرکار خانم مهندس ژاله غفاری و جمعی از دانشجویان کامپیوتر و فناوری اطلاعات دانشگاه کردستان که در جمع آوری داده های مورد نیاز همکاری نمودند، قدردانی می نمایم.

چکیده

امروزه رشد و تغییر نمایی در حجم اطلاعات وب موجب عدم دسترسی سریع و آسان کاربران به اطلاعات مورد نیازشان شده است. اگرچه بکارگیری ابزارهایی مثل 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 پرس‌وجوی کاربر استفاده می‌کنند. در واقع موتورهای جستجو علایق و دانش پیش‌زمینه کاربر را در جستجو مورد توجه قرار نمی‌دهند و نتایج مشابهی را برای کاربران با علایق متفاوت و پرس‌وجوی یکسان بازایی می‌نمایند. بنابراین با توجه به اهمیت دسترسی سریع به اطلاعات مناسب برای کاربران در تعامل با وب، شخصی‌سازی کردن وب به یک پدیده محبوب به منظور سفارشی کردن محیط‌های وب تبدیل شده است. در این میان شخصی‌سازی نتایج حاصل از موتورهای جستجوی وب از اهمیت خاصی برخوردار می‌باشد.

فرآیند شخصی‌سازی می‌تواند بعنوان یک کاربرد داده‌کاوی در نظر گرفته شود که شامل همه فازهای یک چرخه داده‌کاوی رایج از قبیل جمع‌آوری داده، پیش‌پردازش، کشف الگو و توسعه دانش بعنوان واسطه بین کاربر و وب می‌باشد. هدف اصلی در یک سیستم شخصی‌سازی، فراهم نمودن اطلاعات مورد نیاز کاربران برطبق علایق آنها و بدون دخالت صریح کاربران می‌باشد. این در حالی است که بررسی‌های انجام شده نشان می‌دهد که در رویکردهای استفاده شده تاکنون، سیستمی که بتواند فرآیند شخصی‌سازی را بصورت اتوماتیک و مستقل از کاربر و یا شخص خیره انجام دهد، به چشم نمی‌خورد. علاوه بر این در سیستم‌های شخصی‌سازی که بصورت جمعی عمل می‌نمایند، قابلیت بروز رسانی و سازگاری با تغییر در کاربران و علایق آنها وجود ندارد. وجود چنین خلاهایی ما را بر آن داشت تا به دنبال ارائه روش‌های شخصی‌سازی باشیم که علاوه بر اتوماتیک بودن و قابلیت سازگاری با تغییر در کاربران و علایق آنها، از روش‌های هوشمند مبتنی بر یادگیری ماشین مانند تکنیک‌های فازی، الگوریتم‌های داده‌کاوی و غیره در راستای بهبود نتایج بدست آمده و افزایش کارایی سیستم استفاده نماید.

در راستای این اهداف سه روش برای شخصی‌سازی نتایج موتورهای جستجو ارائه می‌شود. در هر سه روش از شبکه‌های مفهومی فازی برای نمایش علایق کاربر استفاده می‌گردد که ابزاری مناسب برای نمایش و پردازش اطلاعات و مفاهیم غیردقیق و فازی مربوط به کاربران می‌باشد.

در روش اول که یک روش شخصی سازی جمعی است، از شبکه های مفهومی فازی اتوماتیک و سازگار در کنار ابزارهای داده کاوی (خوشه بندی) استفاده شده است. در این روش شبکه مفهومی فازی بر اساس پروفایل گروهی از کاربران با علائق مشابه و بصورت اتوماتیک و برون خطی ایجاد می شود. سپس نتایج موتور جستجو برای کاربر جاری با استفاده از شبکه مفهومی فازی متعلق به گروهی که بیشترین شباهت را با کاربر مورد نظر دارند، مجدداً رتبه بندی می شوند.

در روش دوم از شبکه های مفهومی فازی غنی سازی شده با استفاده از مفاهیم آنتولوژی برای شخصی سازی نتایج جستجو استفاده می شود. در این روش رتبه بندی مجدد نتایج جستجوی کاربر با استفاده از شبکه مفهومی فازی غنی سازی شده بر اساس پروفایل کاربر صورت می گیرد.

در روش سوم یک سیستم چندعامله برای شخصی سازی نتایج موتور فراجستجوی شبیه سازی شده، طراحی شده است. از آنجاییکه موتورهای فراجستجو نتایج چندین موتور جستجو را ترکیب می نمایند، قادر هستند برخی از معایب و محدودیت های موتورهای جستجو مانند پوشش جزئی را تا حدودی رفع نمایند. در این روش نیز از شبکه های مفهومی فازی غنی سازی شده برای شخصی سازی نتایج موتور فراجستجوی شبیه سازی شده استفاده می شود. بکارگیری سیستم های چندعامله برای شخصی سازی نتایج موتورهای فراجستجو علاوه بر افزایش کارایی باعث سهولت در توسعه، پشتیبانی و توزیع پذیری آنها خواهد شد.

برای تست و ارزیابی روش های پیشنهادی آزمایشاتی انجام می شوند و دو معیار ارزیابی نیز برای مقایسه کمی روش های پیشنهادی مورد استفاده قرار می گیرد. نتایج بدست آمده و بررسی مزایا و معایب هر یک از این روش ها در هر فصل ارائه شده است.

فهرست مطالب

1	فصل 1. مقدمه
1-1-1	انگیزه تحقیق
2-1-1	بیان مساله تحقیق
3-1-1	اهداف و رویکردهای بکار رفته در تحقیق
4-1-1	دستاوردهای پایان نامه
5-1-1	ساختار پایان نامه
6-1-1	مقالات منتشر شده
8	فصل 2. مطالب پیش زمینه
1-2-1	مقدمه
2-2-1	شخصی سازی وب
1-2-2-1	ضرورت شخصی سازی وب
2-2-2-1	تعریف و اهداف شخصی سازی
3-2-2-1	عملکردهای شخصی سازی وب
1-3-2-2-1	به خاطر سپاری
2-3-2-2-1	راهنمایی
3-3-2-2-1	تطبیق
4-3-2-2-1	پشتیبانی اجرای وظیفه
4-2-2-1	نیازمندی های یک سیستم شخصی سازی وب
5-2-2-1	رویکردهای بکار رفته در شخصی سازی وب
6-2-2-1	آماده سازی و مدلسازی داده
1-6-2-2-1	انواع داده
2-6-2-2-1	آماده سازی داده ها
3-6-2-2-1	مدلسازی داده ها
3-2-3-1	موتورهای جستجو
1-3-2-3-1	طبقه بندی موتورهای جستجو
2-3-2-3-1	بررسی عملکرد موتورهای جستجوی پیمایشی
3-3-2-3-1	نحوه رتبه بندی صفحات وب توسط موتورهای جستجو
4-3-2-3-1	معایب موتورهای موتورهای جستجو

25	4-2- شبکه های مفهومی فازی
26	1-4-2- تعاریف و مفاهیم شبکه های مفهومی فازی
29	2-4-2- پردازش پرس و جو در شبکه مفهومی فازی
31	فصل 3. مروری بر رویکردهای موجود در زمینه شخصی سازی موتورهای جستجو
31	1-3- مقدمه
31	2-3- رویکردهای موجود در زمینه مدلسازی کاربر جهت شخصی سازی موتورهای جستجو
34	3-3- رویکردهای موجود در شخصی سازی نتایج جستجو
35	1-3-3- کارهای انجام شده با استفاده از رویکرد شبکه مفهومی فازی
37	4-3- جمع بندی
38	فصل 4. شخصی سازی جستجو بر مبنای شبکه های مفهومی فازی اتوماتیک و سازگار
38	1-4- مقدمه
38	2-4- ویژگی های رویکرد پیشنهادی
39	3-4- نحوه گردآوری و آماده سازی پروفایل ها
41	4-4- معماری روش پیشنهادی
42	1-4-4- فاز اول
43	2-4-4- فاز دوم
45	5-4- جزئیات پیاده سازی
46	6-4- ارزیابی نتایج
46	1-6-4- پارامترهای ارزیابی
47	2-6-4- آزمایشات انجام شده
47	3-6-4- نتایج آزمایشات
52	4-6-4- تحلیل نتایج آزمایشات
52	7-4- جمع بندی
54	فصل 5. شخصی سازی جستجو با استفاده از شبکه های مفهومی فازی بر اساس آنتولوژی
54	1-5- مقدمه
54	2-5- ویژگی های رویکرد پیشنهادی
56	3-5- معماری روش پیشنهادی
57	1-3-5- مولفه ساخت شبکه مفهومی فازی غنی سازی شده
58	2-3-5- مولفه ایجاد ماتریس توصیف سند غنی سازی شده
59	3-3-5- مولفه رتبه بندی

59	4-5- جزئیات پیاده سازی
59	1-4-5- وارد کردن آنتولوژی
60	2-4-5- پیاده سازی مولفه های سیستم
60	5-5- ارزیابی نتایج
61	1-5-5- آزمایشات انجام شده
61	2-5-5- نتایج آزمایشات
64	3-5-5- تحلیل نتایج آزمایشات
64	6-5- جمع بندی
66	فصل 6. طراحی یک موتور فرا جستجوی شخصی سازی شده بر مبنای سیستم های چندعامله
66	1-6- مقدمه
67	2-6- ویژگی های رویکرد پیشنهادی
67	3-6- عامل ها و سیستم های چندعامله
68	4-6- معماری روش پیشنهادی
68	1-4-6- عامل کاربر
69	2-4-6- گروه عامل های جستجو
70	3-4-6- گروه عامل های شخصی سازی
72	5-6- جزئیات پیاده سازی
72	1-5-6- پیاده سازی مولفه های سیستم
73	6-6- ارزیابی آزمایشات سیستم چندعامله برای شخصی سازی نتایج یک موتور فرا جستجو
73	1-6-6- انجام شده
73	2-6-6- نتایج آزمایشات
77	3-6-6- تحلیل نتایج آزمایشات
77	7-6- مقایسه روش های پیشنهادی
78	8-6- جمع بندی
79	فصل 7. خلاصه، نتیجه گیری و کارهای آینده
79	1-7- مقدمه
79	2-7- خلاصه
83	مراجع
90	پیوست 1: ابزارهای مورد استفاده در پیاده سازی پایان نامه
97	پیوست 2: بردار مفاهیم 300 تایی از پیش تعیین شده

پیوست 3: نتایج ارزیابی روش های پیشنهادی 98

پیوست 4: ارائه مثال های مربوط به محاسبه معیار ارزیابی MAP و محاسبه شبکه مفهومی فازی 109

فهرست اشکال

- شکل 2-1: یک شبکه مفهومی فازی..... 26
- شکل 4-1: قطعه کد بازیابی صفحات وب بازدید شده توسط کاربر..... 41
- شکل 4-2: معماری روش پیشنهادی 1..... 42
- شکل 4-3: معیار ارزیابی میانگین d برای روش پیشنهادی 1 و موتور جستجوی گوگل..... 51
- شکل 4-4: معیار ارزیابی میانگین d برای روش پیشنهادی 1 و موتور جستجوی یاهو..... 51
- شکل 4-5: معیار ارزیابی MAP برای روش پیشنهادی 1 و موتور جستجوی گوگل..... 51
- شکل 4-6: معیار ارزیابی MAP برای روش پیشنهادی 1 و موتور جستجوی یاهو..... 52
- شکل 5-1: معماری روش پیشنهادی 2..... 57
- شکل 5-2: معیار ارزیابی میانگین d برای روش های پیشنهادی 1 و 2 و موتور جستجوی گوگل..... 63
- شکل 5-3: معیار ارزیابی میانگین d برای روش های پیشنهادی 1 و 2 و موتور جستجوی یاهو..... 63
- شکل 5-4: معیار ارزیابی MAP برای روش های پیشنهادی 1 و 2 و موتور جستجوی گوگل..... 64
- شکل 5-5: معیار ارزیابی MAP برای روش های پیشنهادی 1 و 2 و موتور جستجوی یاهو..... 64
- شکل 6-1: معماری روش پیشنهادی..... 68
- شکل 6-2: معماری متداول موتورهای فراجستجو..... 69
- شکل 6-3: معماری 5 سطحی پیشنهادی..... 71
- شکل 6-4: ارتباط بین عامل ها..... 72
- شکل 6-5: معیار ارزیابی میانگین d برای روش پیشنهادی و موتور فراجستجو..... 76
- شکل 6-6: معیار ارزیابی MAP برای روش پیشنهادی و موتور فراجستجو..... 76

فهرست جداول

- جدول 4-1: نتایج موتور جستجوی گوگل..... 48
- جدول 4-2: اسناد مناسب انتخاب شده توسط کاربران..... 48
- جدول 4-3: نتایج جستجوی شخصی سازی شده با استفاده از روش پیشنهادی 1..... 49
- جدول 4-4: میانگین دقت (AP) نتایج به دست آمده با استفاده از روش پیشنهادی 1..... 49
- جدول 4-5: رتبه بندی نتایج موتور جستجوی گوگل توسط کاربران..... 49
- جدول 4-6: رتبه بندی شخصی سازی شده نتایج موتور جستجوی گوگل با استفاده از روش پیشنهادی 1..... 49
- جدول 4-7: معیار ارزیابی d برای موتور جستجوی گوگل و روش پیشنهادی 1 (query="OOP")..... 49
- جدول 4-8: معیار ارزیابی میانگین d برای موتور جستجوی گوگل و روش پیشنهادی 1..... 50
- جدول 4-9: متوسط میانگین دقت (MAP) نتایج به دست آمده با استفاده از روش پیشنهادی 1..... 50
- جدول 4-10: معیار ارزیابی میانگین d برای موتور جستجوی یاهو و روش پیشنهادی 1..... 50
- جدول 4-11: متوسط میانگین دقت (MAP) نتایج به دست آمده با استفاده از روش پیشنهادی 1..... 50
- جدول 5-1: نتایج جستجوی شخصی سازی شده با استفاده از روش پیشنهادی 2..... 61
- جدول 5-2: میانگین دقت (AP) نتایج به دست آمده با استفاده از روش پیشنهادی 2..... 62
- جدول 5-3: رتبه بندی شخصی سازی شده نتایج موتور جستجوی گوگل با استفاده از روش پیشنهادی 2..... 62
- جدول 5-4: معیار ارزیابی d برای موتور جستجوی گوگل و روش پیشنهادی 2 (query="oop")..... 62
- جدول 5-5: معیار ارزیابی میانگین d برای موتور جستجوی گوگل و روش های پیشنهادی 1 و 2..... 62
- جدول 5-6: متوسط میانگین دقت (MAP) نتایج به دست آمده با استفاده از روش پیشنهادی 1 و 2..... 62
- جدول 5-7: معیار ارزیابی میانگین d برای موتور جستجوی یاهو و روش پیشنهادی 1 و 2..... 63
- جدول 5-8: متوسط میانگین دقت (MAP) نتایج به دست آمده با استفاده از روش پیشنهادی 1 و 2..... 63
- جدول 6-1: نتایج موتور فراجستجوی شبیه سازی شده..... 74
- جدول 6-2: اسناد مناسب انتخاب شده توسط کاربران..... 74
- جدول 6-3: نتایج فراجستجوی شخصی سازی شده..... 75
- جدول 6-4: میانگین دقت (AP) نتایج به دست آمده..... 75
- جدول 6-5: رتبه بندی نتایج موتور فراجستجوی شبیه سازی شده توسط کاربران..... 75
- جدول 6-6: رتبه بندی شخصی سازی شده نتایج موتور فراجستجوی شبیه سازی شده..... 75
- جدول 6-7: معیار ارزیابی d برای موتور فراجستجوی شبیه سازی شده و سیستم پیشنهادی..... 75
- جدول 6-8: معیار ارزیابی میانگین d برای موتور فراجستجوی شبیه سازی شده و سیستم پیشنهادی..... 76
- جدول 6-9: متوسط میانگین دقت (MAP) نتایج به دست آمده با استفاده از سیستم پیشنهادی..... 76

جدول 6-10: مقایسه بهبود رتبه‌بندی و دقت نتایج.....78

جدول 6-11: مقایسه زمان مورد نیاز سه روش پیشنهادی.....78

علائم اختصاری و نشانه‌ها

FCN	Fuzzy Concept Network
DDM	Document Descriptor Matrix
EFCN	Enriched Fuzzy Concept Network
EDDM	Enriched Document Descriptor Matrix
SP	Sequential Pattern
CSP	Contiguous Sequential Pattern
MAP	Mean Average Precision
AP	Average Precision
d	distance
tf-idf	term frequency-inverse document frequency

فصل 1. مقدمه

1-1- انگیزه تحقیق

در حال حاضر وب جهانی به منبع عظیمی از اطلاعات که شامل داده‌های ساخت یافته و نیمه ساخت یافته می‌باشد، تبدیل شده که کاربران را با مشکل همپوشانی داده‌ها مواجه کرده است. علاوه بر این بررسی‌های گوناگون انجام شده در زمینه گسترش وب نشان داده است که روزانه بیش از یک میلیون صفحه به وب اضافه می‌شود و بیش از 600 گیگابایت از صفحات آن در هر ماه تغییر می‌کنند [1 و 2]. این رشد و تغییر نمایی در حجم اطلاعات وب سبب بروز مشکلاتی برای کاربران وب شده است. از مهم‌ترین این مشکلات عدم دسترسی سریع و آسان به اطلاعات مورد نیاز می‌باشد. اگرچه موتورهای جستجو توانسته‌اند تا اندازه‌ای این مشکل را هموار نموده و کاربران را در یافتن صفحات وب مرتبط با یک پرس و جو کمک نمایند، اما حجم اطلاعاتی که موتورهای جستجو به کاربر تحویل می‌دهند معمولاً بسیار بیشتر از آن است که قابل پردازش توسط کاربر باشند. بعلاوه بیشتر موتورهای جستجو تنها از محتوای صفحات و ساختار ارتباطی موجود بین آنها برای ارزیابی میزان ارتباط صفحات با پرس و جوی کاربر استفاده می‌کنند. از این رو برای موتورهای جستجو اهمیت ندارد که پرس و جو توسط چه کسی و با چه علایق و دانش پیش‌زمینه‌ای وارد شده‌است و نتایج مشابهی را برای کاربران با علایق متفاوت و پرس و جوی یکسان بازبایی می‌نمایند. بنابراین با توجه به اهمیت دسترسی سریع به اطلاعات مناسب برای کاربران در تعامل با وب، شخصی کردن وب به یک پدیده محبوب به منظور سفارشی کردن محیط‌های وب تبدیل شده است بطوریکه شخصی سازی بعنوان یک فرایند حیاتی که نقش مهمی در موفقیت وب سایت‌ها و مخصوصاً موتورهای جستجو دارد، مطرح می‌گردد.

1-2- بیان مساله تحقیق

امروزه علی‌رغم نیاز کاربران وب به شخصی‌سازی نتایج موتورهای جستجو برای دسترسی به اطلاعات متناسب با علایقشان، هنوز موتور جستجویی که بتواند شخصی‌سازی نتایج جستجو را بصورت کارایی انجام دهد، وجود ندارد. لذا در این پایان نامه به بررسی رویکردهایی برای افزایش کارایی سیستم‌های شخصی‌سازی نتایج جستجو پرداخته می‌شود. با توجه به وسعت و گستردگی زمینه بحث موارد ذیل مورد توجه قرار گرفته است:

1. در سیستم‌های شخصی‌سازی عدم دخالت کاربر در فرآیند شخصی‌سازی از اهمیت ویژه‌ای برخوردار است، این در حالی است که بررسی‌های انجام شده نشان می‌دهد که در رویکردهای موجود سیستمی که بتواند فرآیند شخصی‌سازی را بصورت اتوماتیک و مستقل از کاربر و یا شخص خبره انجام دهد، وجود ندارد.

2. علاوه بر این در سیستم‌های شخصی‌سازی که بصورت جمعی عمل می‌نمایند، عدم وجود قابلیت سازگاری¹ و بروز رسانی با تغییر در کاربران و علایق آنها باعث کاهش کارایی سیستم‌های شخصی‌سازی موجود شده است.

وجود چنین خلاهایی ما را بر آن داشت تا به دنبال ارائه روش‌های شخصی‌سازی باشیم که علاوه بر اتوماتیک بودن و قابلیت سازگاری با تغییر در کاربران و علایق آنها، از روش‌های هوشمند مانند تکنیک‌های فازی، الگوریتم‌های داده‌کاوی²، مفاهیم آنتولوژی و سیستم‌های چندعامله³ در راستای بهبود نتایج بدست آمده و افزایش کارایی سیستم استفاده نماید.

1-3- اهداف و رویکردهای بکار رفته در تحقیق

با توجه به اهمیت شخصی‌سازی موتورهای جستجو در تسهیل دسترسی به اطلاعات وب توسط کاربران و عدم دخالت کاربر در فرآیند شخصی‌سازی در این تحقیق تلاش شده است تا با بکارگیری تکنیک‌های هوشمند و مبتنی بر یادگیری ماشین، نتایج موتورهای جستجو و فرآیند جستجو برای کاربران متفاوت بصورت اتوماتیک و سازگار شخصی‌سازی گردد. در راستای شخصی‌سازی موتورهای جستجو لازم است علایق کاربران در رتبه‌بندی نتایج بازیابی شده مورد توجه قرار گیرد. بنابراین نتایج بازیابی شده توسط موتورهای جستجو برای کاربران متفاوت بر اساس علاقه آنها مجدداً رتبه‌بندی شده و سپس به کاربر ارائه می‌-

¹- Adaptive

²- Data Mining

³- Multi Agent Systems

گردد. بنابراین نیاز به جمع‌آوری اطلاعاتی از کاربر داریم که بیانگر علایق کاربر باشد و در پروفایل شخصی آنها قرار گیرد. در این تحقیق پروفایل کاربر بصورت اتوماتیک و بر اساس تعاملات کاربر با وب و سابقه مرورهای¹ او ایجاد شده است. ما مجموعه‌ای از صفحات وب قبلاً بازدید شده توسط کاربر را جمع‌آوری نموده و در پروفایل کاربر نگهداری می‌کنیم. سپس از پروفایل ایجاد شده برای رتبه‌بندی مجدد نتایج جستجوی کاربر مزبور استفاده می‌نماییم. از آنجاییکه کاربران معمولاً علایق خود را به صورت دقیق و مطمئن بیان نمی‌نمایند و در واقع آنها بیشتر تمایل به بیان علایق خود به صورت فازی و نامطمئن دارند، بنابراین یکی از اهداف اصلی این تحقیق استفاده از روش‌های فازی جهت شخصی‌سازی نتایج موتورهای جستجو می‌باشد. در راستای این هدف ما از شبکه‌های مفهومی فازی² برای نمایش علایق کاربر استفاده نمودیم که ابزاری مناسب برای نمایش و پردازش اطلاعات و مفاهیم غیردقیق و فازی بیان شده توسط کاربران می‌باشد. در شبکه‌های مفهومی فازی استفاده شده در شخصی‌سازی موتورهای جستجو، ارتباط بین مفاهیم توسط فرد خبره تعیین می‌شوند [3 و 4] در حالیکه در این تحقیق تلاش می‌شود تا شبکه مفهومی فازی بصورت اتوماتیک و بر اساس پروفایل کاربر ایجاد گردد. علاوه بر این از ابزارهای داده‌کاوی برای خوشه‌بندی کاربران مشابه و ایجاد شبکه مفهومی فازی برای هر خوشه در یک فاز برون‌خطی³ و در نتیجه افزایش کارایی سیستم استفاده شده است. در راستای بهبود نتایج بدست آمده از مفاهیم آنتولوژی برای غنی‌سازی شبکه‌های مفهومی ایجاد شده استفاده گردید که در واقع گامی در جهت تحقق وب معنایی نیز می‌باشد. بکارگیری معماری چندعامله جهت پیاده‌سازی سیستم پیشنهادی که باعث تسهیل در توسعه، پشتیبانی و توزیع‌پذیری سیستم پیشنهادی و بهبود در کارایی آن می‌شود نیز از اهداف دیگر این تحقیق می‌باشد.

1-4- دستاوردهای پایان‌نامه

در راستای رسیدن به اهداف تحقیق، دستاوردهای مهمی حاصل شد که در ادامه هر یک از این دستاوردها بطور مختصر شرح داده می‌شوند و در فصل‌های بعد به تشریح جزئیات آنها پرداخته خواهد شد. همان‌طوری که در بخش‌های قبلی گفته شد، هدف ارائه روشی است که نتایج موتورهای جستجو را بر اساس سلیقه هر یک از کاربران و با توجه به پروفایل آنها شخصی‌سازی نماید. اولین دستاورد، ارائه یک موتور جستجوی شخصی‌سازی شده با استفاده از شبکه‌های مفهومی فازی اتوماتیک و سازگار در کنار بکارگیری ابزارهای داده‌کاوی است که دارای ویژگی‌های مطلوب زیر می‌باشد:

- گروه‌بندی کاربران با علایق مشابه و ایجاد یک شبکه مفهومی فازی برای هر گروه

¹- Browsing History

²- Fuzzy Concept Networks

³- Offline