



دانشگاه قم

دانشکده فنی و مهندسی

پایان نامه کارشناسی ارشد رشته مهندسی فناوری اطلاعات

عنوان:

طراحی سیستم پیشنهاد دهنده وبلاگ

(مورد مطالعه: پارسای بلاگ)

استاد راهنما:

دکتر بهروز مینایی بیدگلی

استاد مشاور:

مهندس مهدی نصیری

نگارنده:

زینب برهانی فرد

شهریور ماه ۱۳۹۰

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ





دانشگاه قم

دانشکده فنی و مهندسی

پایان نامه کارشناسی ارشد رشته مهندسی فناوری اطلاعات

عنوان:

طراحی سیستم پیشنهاد دهنده وبلاگ

(مورد مطالعه: پارسی بلاگ)

استاد راهنما:

دکتر بهروز مینایی بیدگلی

استاد مشاور:

مهندس مهدی نصیری

نگارنده:

زینب برهانی فرد

شهریور ماه ۱۳۹۰





«صورت جلسه دفاع از پایان نامه کارشناسی ارشد»

**با تائیدات خداوند متعال و با استعانت از حضرت ولی عصر (عجل الله تعالی فرجه الشریف)**

جلسه دفاعیه پایان نامه کارشناسی ارشد آقای/خانم: رشته:

تحت عنوان:

با حضور هیات داوران در محل دانشگاه قم در تاریخ: / / تشکیل گردید.

در این جلسه پایان نامه با نمره (به عدد .....، به حروف.....) و با درجه: عالی

بسیار خوب  خوب  مورد قبول  مورد دفاع قرار گرفت.

نام و نام خانوادگی	سمت	مرتبۀ علمی	امضا
	استاد راهنما		
	استاد مشاور		
	استاد ناظر		
	استاد ناظر		
	نماینده کمیته تحصیلات تکمیلی		

مدیر امور آموزشی و تحصیلات تکمیلی

نام و امضاء

معاون آموزشی و پژوهشی دانشکده

نام و امضاء

## تقدیم به:

...همسر عزیز و مهربانم

... و دختر کوچولو و نازم، سارا

## تشکر و قدردانی:

در اینجا بر خود لازم می‌دانم از استاد ارجمند جناب آقای دکتر مینایی که در طول انجام این پایان نامه با راهنمایی‌های دقیق و موشکافان‌شان مرا یاری نموده‌اند، تقدیر و تشکر نمایم.

از جناب آقای مهندس نصیری به عنوان مشاور که با راهنمایی خود مرا مورد لطف قرار داده‌اند کمال تشکر را دارم.

همچنین از آقای مهندس فخری، مدیریت محترم پارسی بلاگ به علت همکاری صمیمانه در جهت تهیه مجموعه داده تشکر نمایم.

## چکیده

در سال‌های اخیر وبلاگ‌ها تبدیل به رسانه اجتماعی برجسته‌ای بر روی اینترنت شده‌اند که به سرعت کاربران را قادر می‌سازند تا محتوا را بر روی وب منتشر کنند که شامل اندیشه‌های کاملاً شخصی می‌باشد. راحتی انتشار مطالب به همراه میل و اشتیاق برای گسترش افکار، عاملی مهم برای رشد سریع و مداوم وبلاگ‌ها شده است. به دلیل حجم بالای محتوایی که توسط وبلاگ‌ها فراهم می‌شود و با توجه به اینکه اغلب بلاگرها کاربران غیر حرفه‌ای هستند که یافتن وبلاگ‌های همفکر و یا مورد علاقه‌شان سخت می‌باشد. بنابراین با توجه به انبوه اطلاعات برای وبلاگ‌ها، وجود یک سیستم پیشنهادگر کاملاً ضروری به نظر می‌رسد.

سیستم‌های پیشنهادگر از نظرات گروهی از کاربران استفاده می‌کنند تا اینکه به طور مؤثرتر به افراد در آن گروه کمک کنند تا اقلام مورد علاقه‌شان را از یک مجموعه بسیار عظیم از انتخاب‌ها شناسایی کنند. سیستم‌های پیشنهادگر دارای تکنیک‌های متنوعی هستند و از منابع داده‌ای مختلف برای ارائه پیشنهاد به کاربران استفاده می‌کنند.

سیستم‌های پیشنهادگر برای وبلاگ‌ها بر مبنای اطلاعات محتوای وبلاگ‌ها، شبکه اجتماعی و پروفایل آن‌ها طراحی می‌شوند. در این پروژه، روش‌های مختلف در سیستم‌های پیشنهادگر وبلاگ‌ها بررسی و ارزیابی می‌شود همچنین یک چارچوب پیشنهاد وبلاگ را برای تولید پیشنهادها ارائه می‌دهیم.

**کلمات کلیدی:** سیستم پیشنهادگر، وبلاگ، شبکه اجتماعی، خوشه‌بندی، گراف

# فهرست مطالب

صفحه	عنوان
۱	فصل ۱. مقدمه.....
۲	۱-۱. مقدمه.....
۳	۲-۱. تعریف مسئله و بیان موضوع اصلی تحقیق.....
۴	۳-۱. ضرورت و اهمیت تحقیق.....
۶	۴-۱. هدف تحقیق.....
۶	۵-۱. فرضیه‌های تحقیق.....
۷	۶-۱. کارهای مرتبط با سیستم‌های پیشنهاد گر وبلاگ.....
۹	۷-۱. مجموعه داده مورد استفاده در پروژه.....
۱۰	۸-۱. ساختار پایان‌نامه.....
۱۱	فصل ۲. تحلیل شبکه اجتماعی.....
۱۲	۱-۲. مقدمه.....
۱۴	۲-۲. تحلیل شبکه اجتماعی.....
۱۵	۳-۲. روش‌های رسمی برای نمایش شبکه‌های اجتماعی.....
۱۶	۱-۳-۲. استفاده از گراف برای نمایش روابط اجتماعی.....
۱۸	۲-۳-۲. استفاده از ماتریس برای نمایش روابط اجتماعی.....
۱۹	۳-۳-۲. روش‌های آماری برای تحلیل شبکه‌های اجتماعی.....
۱۹	۴-۲. خصوصیات شبکه‌های اجتماعی.....
۲۰	۱-۴-۲. اندازه شبکه.....
۲۰	۲-۴-۲. چگالی شبکه.....
۲۰	۳-۴-۲. درجه بازیگران.....
۲۱	۴-۴-۲. حداکثر شار.....
۲۱	۵-۴-۲. همبستگی هابل و کاتز.....
۲۲	۶-۴-۲. اثر تیلور.....
۲۲	۷-۴-۲. مرکزیت و قدرت.....

۲۳	..... ۲-۴-۷-۱. درجه
۲۳	..... ۲-۴-۷-۲. نزدیکی
۲۴	..... ۲-۴-۷-۳. بینایی
۲۵	..... ۲-۴-۸. کانون و مرجع
۲۵	..... ۲-۴-۹. ضریب تأثیر خوشه‌بندی
۲۶	..... ۲-۴-۱۰. دسترس پذیری
۲۶	..... ۲-۴-۱۱. مؤلفه همبند قوی و ضعیف
۲۷	..... ۲-۵-۵. خوشه بندی شبکه
۲۸	..... ۲-۵-۱. الگوریتم حریرانه سریع
۲۹	..... ۲-۵-۲. الگوریتم بردار ویژه پیشرو
۳۰	..... ۲-۶-۶. الگوریتم‌های رتبه بندی بر اساس تحلیل پیوند
۳۰	..... ۲-۶-۱. الگوریتم PageRank
۳۱	..... ۲-۶-۲. الگوریتم HITS
۳۲	..... ۲-۷-۷. شناخت فضا و شبکه اجتماعی وبلاگ
۳۳	..... ۲-۷-۱. شناخت اجزای وبلاگ
۳۴	..... ۲-۷-۲. پیوندها در فضای وبلاگ
۳۶	..... ۲-۸. ابزارهای تحلیل شبکه‌های اجتماعی
۳۷	..... ۲-۹. وب کاوی
۳۹	..... ۲-۱۰. جمع بندی فصل
۴۰	..... <b>فصل ۳. سیستم‌های پیشنهادگر</b>
۴۱	..... ۳-۱. مقدمه
۴۱	..... ۳-۲. مفهوم سیستم پیشنهادگر
۴۳	..... ۳-۳. تکنیک‌های مختلف سیستم‌های پیشنهادگر
۴۵	..... ۳-۳-۱. سیستم پیشنهادگر مبتنی بر فیلترینگ مشارکتی
۴۶	..... ۳-۳-۱-۱. روش‌های فیلترینگ مشارکتی مبتنی بر حافظه
۴۷	..... ۳-۳-۱-۱-۱. محاسبه شباهت
۵۰	..... ۳-۳-۱-۱-۲. پیش بینی و محاسبه پیشنهادات
۵۱	..... ۳-۳-۱-۱-۲. پیشنهاد N-بالا
۵۳	..... ۳-۳-۱-۲. روش‌های فیلترینگ مشارکتی مبتنی بر مدل
۵۳	..... ۳-۳-۲. سیستم پیشنهادگر مبتنی بر محتوا

۵۴	..... سیستم پیشنهادگر مبتنی بر دانش..... ۳-۳-۳
۵۵	..... سیستم پیشنهادگر مبتنی بر جمعیت شناختی ..... ۴-۳-۳
۵۵	..... سیستم پیشنهادگر مبتنی بر مطلوبیت..... ۵-۳-۳
۵۶	..... سیستم پیشنهادگر ترکیبی..... ۶-۳-۳
۵۷	..... سیستم پیشنهادگر در شبکه‌های اجتماعی..... ۴-۳
۵۸	..... سیستم پیشنهادگر در فضای وبلاگ..... ۵-۳
۶۱	..... جمع‌بندی فصل..... ۶-۳
<b>۶۲</b>	<b>..... فصل ۴. پیش-پردازش و تحلیل مجموعه داده..... ۶۲</b>
۶۳	..... ۱-۴. مقدمه..... ۶۳
۶۳	..... ۲-۴. پیش-پردازش..... ۶۳
۶۴	..... ۱-۲-۴. پیش پردازش داده‌های محتوایی..... ۶۴
۶۷	..... ۱-۱-۲-۴. حذف برچسب‌های HTML و یکسان سازی متن..... ۶۷
۶۸	..... ۲-۱-۲-۴. استخراج کلمات..... ۶۸
۷۱	..... ۳-۱-۲-۴. وزن‌دهی و تشکیل فضای بردار واژه‌ها..... ۷۱
۷۲	..... ۴-۱-۲-۴. محاسبه میزان مشابهت میان بلاگ‌ها..... ۷۲
۷۲	..... ۲-۲-۴. پیش پردازش شبکه اجتماعی وبلاگ‌ها..... ۷۲
۷۶	..... ۳-۴. تحلیل آماری مجموعه داده..... ۷۶
۷۶	..... ۱-۳-۴. خصوصیات گراف بلاگ..... ۷۶
۷۸	..... ۲-۳-۴. تحلیل آماری پروفایل وبلاگ‌ها..... ۷۸
۸۰	..... ۳-۳-۴. تحلیل پست‌ها و نظرات..... ۸۰
۸۲	..... ۴-۴. رتبه بندی وبلاگ‌ها..... ۸۲
۸۲	..... ۱-۴-۴. رتبه بندی وبلاگ‌ها بر حسب تعداد پست‌ها و تعداد نظرات..... ۸۲
۸۵	..... ۲-۴-۴. رتبه بندی وبلاگ‌ها بر حسب تعداد لینک ورودی..... ۸۵
۸۷	..... ۳-۴-۴. رتبه بندی بر مبنای الگوریتم HITS..... ۸۷
۸۸	..... ۴-۴-۴. رتبه بندی بر مبنای الگوریتم PageRank..... ۸۸
۸۹	..... ۵-۴. جمع بندی فصل..... ۸۹
<b>۹۰</b>	<b>..... فصل ۵. سیستم پیشنهادگر وبلاگ و نتایج تجربی..... ۹۰</b>
۹۱	..... ۲-۵. مقدمه..... ۹۱
۹۱	..... ۳-۵. پیشنهادات غیر شخصی..... ۹۱
۹۲	..... ۴-۵. پیشنهاد غیر شخصی..... ۹۲

۹۲	۱-۴-۵. پیشنهاد فیلترینگ مشارکتی مبتنی بر حافظه (kNN-CF).....
۹۳	۲-۴-۵. پیشنهاد مبتنی بر محتوا (kNN-CB).....
۹۳	۳-۴-۵. پیشنهاد مبتنی بر الگوریتم FOAF.....
۹۴	۴-۴-۵. چارچوب پیشنهاد وبلاگ مبتنی بر خوشه‌بندی.....
۹۵	۵-۵. آماده سازی داده.....
۹۵	۱-۵-۵. خوشه بندی.....
۹۸	۲-۵-۵. پست پردازش.....
۹۹	۶-۵. پیاده سازی مدل.....
۹۹	۱-۶-۵. ساخت مدل.....
۱۰۰	۲-۶-۵. تولید پیشنهادات.....
۱۰۱	۷-۵. ارزیابی تجربی.....
۱۰۱	۱-۷-۵. معیارهای ارزیابی.....
۱۰۲	۲-۷-۵. نتایج تجربی.....
۱۰۶	۸-۵. جمع‌بندی فصل.....
۱۰۷	<b>فصل ۶. نتیجه‌گیری و کارهای آتی.....</b>
۱۰۸	۱-۶. جمع‌بندی.....
۱۰۹	۲-۶. پیشنهادات برای کارهای آتی.....
۱۱۱	<b>مراجع.....</b>
۱۱۴	<b>پیوست.....</b>
۱۱۵	نتایج خوشه‌بندی با الگوریتم حریمانه سریع.....
۱۱۵	نتایج خوشه‌بندی با الگوریتم بردار ویژه پیشرو.....
۱۱۵	نتایج خوشه‌بندی با الگوریتم MCODE.....
۱۱۶	نتایج خوشه‌بندی با الگوریتم AP.....



# فهرست شکل‌ها

عنوان	صفحه
شکل ۱-۲. فراوانی کلمه کلیدی "شبکه اجتماعی" در چکیده مقالات در طی سال‌های مختلف.....	۱۵
شکل ۲-۲. استفاده از گراف برای نمایش ساختار اجتماعی.....	۱۷
شکل ۳-۲. نمونه‌ای از محاسبه نزدیکی در میان گره‌های شبکه.....	۲۴
شکل ۴-۲. شبکه غیر جهت‌دار که در آن گره D و I بالاترین درجه بینایی را دارند.....	۲۵
شکل ۵-۲. انواع پیوند در شبکه وبلاگ.....	۳۶
شکل ۱-۳. تکنیک‌های پیشنهادگر مختلف و منابع دانش آن‌ها.....	۴۵
شکل ۲-۳. محاسبه مشابهت مبتنی بر آیت $w_i, j$ .....	۵۰
شکل ۱-۴. مراحل پیش پردازش داده‌های محتوایی.....	۶۷
شکل ۲-۴. شبکه اجتماعی پارسی بلاگ پیش-پردازش شده.....	۷۴
شکل ۳-۴. توزیع مؤلفه همبند قوی در شبکه پارسی بلاگ.....	۷۵
شکل ۴-۴. فراوانی تعداد وبلاگ‌ها بر مبنای پیوند ورودی.....	۸۶
شکل ۱-۵. چارچوب سیستم پیشنهادگر وبلاگ.....	۹۵
شکل ۲-۵. خوشه‌بندی گراف پارسی بلاگ بر مبنای الگوریتم حریرانه سریع.....	۹۷
شکل ۳-۵. خوشه‌بندی گراف شبکه پارسی بلاگ بر مبنای الگوریتم بردار ویژه پیشرو.....	۹۷
شکل ۱ پیوست. خوشه‌بندی گراف پارسی بلاگ با الگوریتم MCODE.....	۱۱۶
شکل ۲ پیوست. خوشه‌بندی گراف پارسی بلاگ با الگوریتم AP.....	۱۱۷

# فهرست جدول‌ها

عنوان	صفحه
جدول ۱-۲. رشد شبکه‌های اجتماعی به تفکیک قاره برای مخاطبان بیشتر از ۱۵ سال (مقایسه ۲۰۰۸ با ۲۰۰۷).....	۱۳
جدول ۲-۲. استفاده از ماتریس برای نمایش شبکه اجتماعی.....	۱۸
جدول ۱-۳. انواع تکنیک‌های مختلف پیشنهاد دهنده.....	۴۴
جدول ۲-۳. روش‌های ترکیب تکنیک‌های پیشنهادگر.....	۵۶
جدول ۱-۴. صورت‌های مختلف پیشوند و پسوند در کلمات.....	۶۸
جدول ۲-۴. حالت‌های مختلف املائی کلمه.....	۶۸
جدول ۳-۴. لیست کلمات بی‌فایده در زبان فارسی.....	۷۰
جدول ۴-۴. مقایسه ویژگی‌های شبکه‌های مختلف در پارسی بلاگ.....	۷۴
جدول ۵-۴. مقایسه اطلاعات شبکه قبل و بعد از انجام پیش-پردازش.....	۷۵
جدول ۶-۴. مؤلفه‌های همبند قوی در شبکه پارسی بلاگ.....	۷۶
جدول ۷-۴. مؤلفه‌های همبند ضعیف در شبکه پارسی بلاگ.....	۷۷
جدول ۸-۴. خصوصیات شبکه اجتماعی پارسی بلاگ.....	۷۸
جدول ۹-۴. رتبه بندی وبلاگ‌ها بر حسب تعداد کل پست‌ها.....	۸۳
جدول ۱۰-۴. رتبه بندی وبلاگ‌ها بر حسب تعداد کل پست‌های سال ۸۹.....	۸۴
جدول ۱۱-۴. رتبه بندی وبلاگ‌ها بر مبنای میانگین تعداد نظرات برای وبلاگ سال ۸۹.....	۸۴
جدول ۱۲-۴. لیست ده پست برتر دارای بیشترین تعداد نظرات.....	۸۵
جدول ۱۳-۴. فهرست وبلاگ‌های دارای بالاترین رتبه بر مبنای تعداد پیوندهای ورودی.....	۸۶
جدول ۱۴-۴. لیست وبلاگ‌های با بالاترین مقدار کانون.....	۸۷
جدول ۱۵-۴. لیست وبلاگ‌های با بالاترین میزان مرجع.....	۸۸
جدول ۱۶-۴. لیست وبلاگ‌های با بالاترین مقدار PageRank.....	۸۸
جدول ۱-۵. مقایسه ویژگی‌های شبکه پارسی بلاگ قبل و بعد از پیش-پردازش و پست پردازش.....	۹۸
جدول ۲-۵. مقادیر دقت و بازخوانی در روش‌های مختلف پیشنهاد.....	۱۰۴

## فهرست نمودارها

<u>صفحه</u>	<u>عنوان</u>
۷۹.....	نمودار ۱-۴ مقایسه وبلاگ‌ها بر اساس جنسیت
۷۹.....	نمودار ۲-۴. مقایسه وبلاگ‌ها بر اساس محدوده سنی
۸۰.....	نمودار ۳-۴. مقایسه وبلاگ‌ها بر اساس سطح تحصیلات
۸۰.....	نمودار ۴-۴. مقایسه وبلاگ‌ها بر اساس وضعیت تأهل
۸۱.....	نمودار ۵-۴. فراوانی وبلاگ‌ها در ساعات مختلف شبانه روز
۸۲.....	نمودار ۶-۴. فراوانی وبلاگ‌ها بر اساس ماه‌های مختلف سال
۸۲.....	نمودار ۷-۴. فراوانی وبلاگ‌ها بر اساس میانگین تعداد نظرات کل برای هر وبلاگ
۹۶.....	نمودار ۱-۵. توزیع اندازه خوشه‌ها و توزیع اندازه مؤلفه همبند قوی
۱۰۵.....	نمودار ۲-۵. مقایسه روش‌های پیشنهاد غیر شخصی
۱۰۵.....	نمودار ۳-۵. مقایسه مقادیر دقت و بازخوانی در روش‌های مختلف الگوریتم FOAF
۱۰۵.....	نمودار ۴-۵. مقایسه روش‌های پیشنهاد در چارچوب وبلاگ

# فهرست فرمول‌ها

<u>عنوان</u>	<u>صفحه</u>
فرمول ۱-۲. ضریب تأثیر خوشه‌بندی در گراف جهت‌دار برای هر گره	۲۶
فرمول ۲-۲. دسترس‌پذیری برای هر گره در گراف	۲۶
فرمول ۳-۲. محاسبه معیار پیمانان	۲۸
فرمول ۴-۲. محاسبه رتبه pagerank برای یک صفحه	۳۱
فرمول ۵-۲. محاسبه مقادیر کانون و مرجع در الگوریتم HITS	۳۱
فرمول ۱-۳. محاسبه مشابهت کسینوسی	۴۸
فرمول ۲-۳. محاسبه مشابهت کسینوسی میان دو بردار	۴۸
فرمول ۳-۳. همبستگی پیرسون بین دو کاربر	۴۹
فرمول ۴-۳. محاسبه همبستگی پیرسون میان دو آیتم	۴۹
فرمول ۵-۳. محاسبه میانگین توزین شده کاربر a بر روی آیتم i	۵۱
فرمول ۶-۳. محاسبه میانگین توزین شده کاربر u بر روی آیتم i	۵۱
فرمول ۱-۴. محاسبه وزن کلمات بر اساس معیار TF/IDF	۷۱
فرمول ۱-۵. ماتریس مجاورت	۹۹
فرمول ۲-۵. محاسبه دقت	۱۰۲
فرمول ۳-۵. محاسبه بازخوانی	۱۰۲

## فصل ۱. مقدمه

## ۱-۱. مقدمه

با رشد سریع شبکه اینترنت و گسترش اطلاعات موجود در آن، حجم اطلاعات موجود در وب بیشتر از حد توانایی کاربران در مدیریت، جذب و به روز نگه‌داری اطلاعات است. یکی از راه حل‌های این مشکل، سیستم‌های شخصی‌ساز<sup>۱</sup> و پیشنهادگر هستند. سیستم‌های پیشنهادگر<sup>۲</sup> یا از نظرات گروهی از کاربران استفاده می‌کنند تا اینکه به طور مؤثرتر به افراد در آن گروه کمک کنند تا اقلام<sup>۳</sup> مورد علاقه‌شان را از یک مجموعه بسیار عظیم از انتخاب‌ها شناسایی کنند.

در سال‌های اخیر وبلاگ‌های تبدیل به رسانه اجتماعی<sup>۴</sup> برجسته‌ای بر روی اینترنت شده‌اند که به سرعت کاربران را قادر می‌سازند تا محتوا را بر روی وب منتشر کنند که شامل اندیشه‌های کاملاً شخصی می‌باشد. راحتی انتشار مطالب به همراه میل و اشتیاق برای گسترش افکار، عاملی مهم برای رشد سریع و مداوم وبلاگ‌ها شده است. در حال حاضر بیش از صدها میلیون وبلاگ در سراسر جهان وجود دارند که به سرعت بر تعداد آن‌ها افزوده می‌شود. از آنجا که هر وبلاگ می‌تواند پیوندهایی به صفحات وب و یا سایر وبلاگ‌های دیگر داشته باشد، ساختار پیوند<sup>۵</sup> وبلاگ‌ها می‌تواند به عنوان یک شبکه اجتماعی<sup>۶</sup> در نظر گرفته شود.

در طی سال‌های اخیر وب رشد بسیار زیادی را در استفاده از وبلاگ‌ها مشاهده کرده است. با توجه به حجم اطلاعات بالا در وبلاگ‌ها، بلاگر<sup>۷</sup>های برای یافتن اطلاعات مورد نظر و پیدا کردن

---

<sup>۱</sup>Personalized systems

<sup>۲</sup>Recommender system

<sup>۳</sup>item

<sup>۴</sup>weblog

<sup>۵</sup>Social media

<sup>۶</sup>Link structure

<sup>۷</sup>Social network

<sup>۸</sup>blogger

وبلاگ‌هایی با افکار و تمایلات مشابه، دچار مشکل هستند. بنابراین با توجه به انبوه اطلاعات برای وبلاگ‌ها، وجود یک سیستم پیشنهادگر کاملاً ضروری به نظر می‌رسد.

در این فصل کلیات تحقیق ارائه می‌شود. در ابتدا مسئله مورد بررسی و موضوع اصلی تحقیق را تعریف کرده و به بیان ضرورت، اهداف و سوالات تحقیق می‌پردازیم. پس از آن به صورت خلاصه روش تحقیق را مرور کرده و قلمرو تحقیق، سابقه تحقیقات موجود، محدودیت‌ها و کاربردهای تحقیق را بیان می‌کنیم. نهایتاً هم ساختار کلی پایان‌نامه ارائه می‌گردد.

## ۱-۲. تعریف مسئله و بیان موضوع اصلی تحقیق

در اثر پیشرفت سریع فناوری اطلاعات و ارتباطات، شبکه جهانی اینترنت دچار افزایش انفجاری حجم مستندات و داده‌های دیجیتال شده است. در این عصر، به علت سادگی امکان انتشار مطالب در وب و در دسترس بودن اینترنت، انسان‌ها بیش از پیش از این فناوری استفاده کرده، به استفاده، تولید و نشر اطلاعات می‌پردازند. این افزایش منابع اطلاعاتی، تعداد انتخاب‌های ممکن هر فرد را برای یافتن منبع مورد نظر بیشتر می‌کند. به دلیل وجود انتخاب‌های متفاوت با کیفیت‌های مختلف برای هر مطلب، ساماندهی نامرتب اطلاعات موجود در وب، تغییرات سریع این اطلاعات و کمبود زمان لازم برای بررسی این منابع، مخصوصاً از نظر میزان صحت اطلاعات هر منبع، تصمیم‌گیری افراد برای گزینش بهترین مرجع با مشکلات بسیاری همراه شده است. در این حالت، حجم اطلاعات در دسترس کاربران وب در حدی زیاد است که قابلیت تصمیم‌گیری و یا بروز نگهداری اطلاعات، راجع به یک موضوع خاص، از کاربران سلب می‌شود. این مشکل، که ناشی از پیشرفت فناوری اطلاعات است، به گرانباری اطلاعات<sup>۱</sup> یا سربار اطلاعات معروف است. یکی از راه‌حل‌ها برای مسئله گرانباری اطلاعات، روش‌های شخصی سازی<sup>۲</sup> است. سیستم‌های شخصی سازی که از یک سیستم پیشنهادگر بهره می‌برند، سعی بر تنظیم پویای صفحات وب بر اساس علائق شخصی کاربران دارند.

---

<sup>۱</sup> Information overload

<sup>۲</sup> Personalization

با ظهور نسل جدید وب موسوم به وب ۲.۰ که وبلاگ‌ها و ویکی‌ها<sup>۱</sup> اجزای اصلی آن هستند، تحلیل وبلاگ‌ها اهمیت بسیار بیشتری یافته است. یک وبلاگ، معمولاً یک وب سایت شامل ورودی‌های تاریخ‌دار (که معمولاً پست نامیده می‌شود) با ترتیب معکوس زمانی می‌باشد و توسط یک کاربر که بلاگر<sup>۲</sup> نامیده می‌شود با استفاده از یک ابزار خاص نوشته و نگهداری می‌شود. موضوع یک وبلاگ می‌تواند بسیار متغیر باشد، از لینک و توضیحات راجع به وبسایت‌های دیگر تا عکس، فیلم، روز نوشت و اخبار شرکت‌ها و افراد. در حال حاضر وبلاگ‌ها تبدیل به ابزاری جهت انتشار سریع ایده‌ها و اطلاعات بر روی وب شده است.

مقدار زیادی اطلاعات در فضای وبلاگی شامل یادداشت‌های متنی (پست‌ها و اظهار نظرات کاربران)، پروفایل‌ها، تصاویر، چند رسانه‌ای و غیره وجود دارد که مرتباً در طول زمان رو به افزایش می‌باشد. با توجه به حجم اطلاعاتی بالا در وبلاگ‌ها و با توجه به اینکه اغلب بلاگرها کاربران غیر حرفه‌ای هستند، خوانندگان وبلاگ‌ها برای یافتن اطلاعات مورد نظر و همچنین بلاگرها برای پیدا کردن وبلاگ‌هایی که دارای افکار و تمایلات مشابه با آن‌ها هستند، دچار مشکل هستند و یا اینکه این عملیات برای آن‌ها وقت‌گیر و غیر بهینه است.

بنابراین با توجه به انبوه اطلاعات در فضای وبلاگ، وجود یک سیستم پیشنهادگر برای وبلاگ‌ها کاملاً ضروری به نظر می‌رسد. این سیستم‌ها سعی بر ارائه اقلام اطلاعاتی کاربر و به عبارت دیگر شخصی سازی تجربه کاربر از وب دارند.

### ۱-۳. ضرورت و اهمیت تحقیق

اطلاعات مفید بسیار زیادی به صورت متن، تصویر، صوت و غیره در وبلاگ‌ها وجود دارد. با توجه به اینکه وبلاگ‌ها مرتباً در حال افزایش هستند و اکثر وبلاگ‌های قبلی هم مرتباً به روزرسانی می‌شوند، بنابراین در سطح وبلاگ‌ها با حجم انبوهی از اطلاعات مواجه هستیم. علاوه بر این، بلاگرها علاقه زیادی به شناسایی بلاگرهای دیگر با تمایلات و علایق مشابه خودشان دارند. در میان تعداد بسیار زیادی گزینه‌های متعدد که در برابر خوانندگان بلاگ‌ها و بلاگرها وجود دارد،

---

<sup>۱</sup> Wiki

<sup>۲</sup> Blogger