

## بهبود عملکرد الگوریتم رتبه بندی تحلیل پیوند HITS با استفاده از چارچوب نگاشت کاهش



فضیلت حججی<sup>۱</sup>، محمد علی نعمت بخش<sup>۲</sup>

<sup>۱</sup> دانشجوی دکتری مهندسی نرم افزار، دانشگاه اصفهان  
<sup>۲</sup> عضو هیئت عملی دانشکده مهندسی کامپیوتر، دانشگاه اصفهان

[mnematbakhsh@yahoo.com](mailto:mnematbakhsh@yahoo.com)

### خلاصه

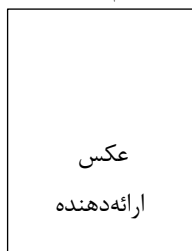
با گسترش روز افزون شبکه جهانی اینترنت و افزایش حجم اطلاعات، نیاز به توسعه روش‌ها و الگوریتم‌های هوشمند برای بازیابی اطلاعات افزایش یافته است. رتبه بندی همواره یکی از بخش‌های مهم در هر سیستم بازیابی اطلاعات است. با توجه به رشد وسیع اطلاعات نیاز به توابع سریع رتبه بندی برای پوشش نیازهای کاربران الزامی است. در این مقاله روشی برای رتبه بندی اسناد مبتنی بر الگوریتم های تحلیل پیوند برای دستیابی به سرعت و کارایی بالاتر ارائه می شود. در روش پیشنهادی از چارچوب نگاشت کاهش برای پیاده سازی الگوریتم استفاده شده است. الگوریتم پیشنهادی به دلیل توزیع بار محاسبات و پیچیدگی های اجرایی الگوریتم بر روی مجموعه ای از خوشه ها تاثیر بسزایی در افزایش سرعت رتبه بندی اسناد دارد.

کلمات کلیدی: بازیابی اطلاعات، رتبه بندی، تحلیل پیوند، نگاشت کاهش

### ۱. مقدمه

گسترش اینترنت باعث افزایش حجم اطلاعات و اسناد الکترونیکی در وب شده است. دسترسی سریع و دقیق به اطلاعات در وب نیازمند به کارگیری روش های هوشمند جهت نزدیکتر شدن فرایند جستجوی اینترنت به نیازهای واقعی کاربران می باشد. بنابراین آنچه امروزه از اهمیت زیادی برخوردار است به کارگیری روش ها و الگوریتم ها در جهت بازیابی و بهره برداری اطلاعات موجود به نحو بهینه است [۱]. رتبه بندی اسناد بخش مهمی از مسائل بازیابی اطلاعات است. حدود ۸۰ درصد از کاربران اینترنت را از طریق موتورهای جستجو کشف و بازدید می کنند. [۲] به علاوه تعداد کاربرانی که از موتورهای جستجو استفاده می کنند به نسبت محتوا در حال افزایش است. عموماً کاربران انتظار دارند که پرس و جوی مورد نظر آنها در یک زمان کوتاه انجام گردد. بنابراین داشتن ابزار مناسب برای جستجوی موثر در وب یکی از مسائل مهم کاربران است. بخش رتبه بندی یکی از مهمترین قسمت های موتور جستجو می باشد. "رتبه بندی" فرآیندی است که کیفیت یک صفحه توسط موتور جستجو تخمین زده می شود [۱].

یک موتور جستجو از اجزای مختلفی نظیر پیمایش گر یا خزنده<sup>۱</sup>، فهرست گر<sup>۲</sup> و تحلیل گر پرس و جو<sup>۳</sup> تشکیل شده است. وظیفه قسمت تحلیل گر پرس و جو دریافت پرس و جوهای کاربران و بازگرداندن صفحات وب متناسب با پرس و جو است [۲]. بعد از استخراج صفحات متناسب با استفاده از الگوریتم رتبه دهی پاسخها بر اساس میزان تناسب و اهمیت مرتب می شوند تا کاربران به صفحات با کیفیت سریع تر دسترسی یابند. الگوریتم رتبه بندی یک موتور جستجو از



۱ Crawler  
 ۲ Indexer  
 ۳ Query Analyzer