

## طبقه‌بندی موضوعی خودکار مقالات فارسی در وب

حبیب‌اله عبدالحسین پورفرد<sup>۲۰۳</sup>، غلامحسین مزدورانی شیرازی<sup>۲۰۴</sup>

### چکیده

صفحات وب با سرعت زیادی در حال گسترش هستند ولی تمام اسناد موجود در وب، از جمله مقالات علمی بصورت منطقی سازماندهی نشده‌اند و این مورد باعث شده تا استفاده و بازیابی آنها با مشکل مواجه شود. طبقه‌بندی موضوعی صفحات وب و به ویژه مقالات علمی دسترسی به اطلاعات را بسیار آسان‌تر و سریع‌تر و دقیق‌تر می‌سازد. در حال حاضر طبقه‌بندی موضوعی مقالات فارسی به صورت پراکنده، محدود و به صورت دستی و غیرخودکار صورت می‌گیرد که با توجه به وفور اطلاعات و رشد روز افزون، با مشکل مواجه است. به خاطر ذات صفحات وب که به صورت نیمه ساخت‌یافته هستند می‌توان اطلاعاتی از ساختار صفحات وب استخراج کرد که به روش‌های معمول طبقه‌بندی متون کمک می‌کند. در این مقاله روشی پیشنهاد شده که الگویی برای تشخیص مقالات فارسی در وب به وسیله ربات خزنده ارائه می‌دهد و در گام بعدی با کمک وب کاوی ساختار رابطه موضوعی بین صفحات وب کشف و از آن در حدس زدن بهتر موضوع مقاله بهره گرفته می‌شود.

### کلمات کلیدی

وب، طبقه‌بندی متون، کشف موضوع، ربات خزنده وب، وب‌کاوی ساختار

<sup>۲۰۳</sup> دانشجوی کارشناسی ارشد کامپیوتر، دانشگاه آزاد اسلامی، واحد تهران جنوب، pourfard@gmail.com

<sup>۲۰۴</sup> عضو هیئت علمی دانشگاه صنعتی مالک اشتر، دانشکده فناوری اطلاعات، shirazi@mut.ac.ir