



دومین کنفرانس متمرکز سیستم های فازی و هوشمند
دانشگاه صنعتی مالک اشتر (پریمیس تهران)
۱۴۰۷ آبان ماه ۹۶



پیمايش موضوعي وب با استفاده از اتوماتاي يادگير سلولي ناهمگام

محمد رضا ميدى، دانشکده مهندسي كامپيوتر و فناوري اطلاعات دانشگاه صنعتي اميركبير
سارا مطيعي، دانشکده مهندسي كامپيوتر و فناوري اطلاعات دانشگاه صنعتي اميركبير

چکیده:

پیمايش وب برای جمع آوری صفحاتی به کار می رود که توسط موتور جستجو شناخت گذاری می شوند. موتور جستجو برای پاسخ دادن به پرس و جوی کاربران از این صفحات پیمايش شده و شناخت گذاری شده استفاده می کند. از آن جا که حجم صفحات وب بسیار بالا و همواره رو به افزایش است، موتورهای جستجو قادر به شناخت گذاری صفحات محدودی هستند. به همین جهت در سال های اخیر به جای پیمايش کل وب، الگوریتم های پیمايش موضوعي یا متصرکز مورد توجه محققین قرار گرفته است که در حین پیمايش به صورت انتخاب گر عمل نموده و صفحات مرتبط با یک موضوع خاص را جمع آوری می کنند. در این مقاله روشی مبتنی بر اتوماتای يادگير سلولي ناهمگام برای پیمايش موضوعي صفحات وب پیشنهاد می گردد. روش پیشنهادی که در آن از ترکیب تکنیک های کاوش ساختار وب و کاوش استفاده از وب استفاده شده، شامل دو مرحله کلی می باشد. در مرحله اول، با استفاده از اتوماتای يادگير سلولي ناهمگام، رفتار کاربران در مشاهده صفحات وب و پیوندهای بین صفحات، ساختار ارتباطي صفحات وب به دست می آید. به آن معنی که صفحات مرتبط با یکدیگر و میزان ارتباط آنها تعیین می شود. در مرحله دوم، با استفاده از ساختار ارتباطي به دست آمده از مرحله قبل، پیمايش موضوعي وب انجام می شود و صفحات مرتبط با یک موضوع خاص جمع آوری می گردد. نتایج ارزیابی پیمايشگر پیشنهادی حاکی از بهبود عملکرد آن (نرخ حاصل و فراخوان هدف) نسبت به پیمايشگر Best First و عدم وابستگی آن به انتخاب مجموعه آغازین دارد.