



مومین غنکزه مشترک سیستم های فازی و هوشمند
دانشگاه صنعتی مالک اشتر (پردیس تهران)
۱۳۸۷ آبان ماه



پیمایش موضوعی وب با استفاده از اتوماتای یادگیر سلولی ناهمگام

محمدرضا میبیدی، دانشکده مهندسی کامپیوتر و فناوری اطلاعات دانشگاه صنعتی امیرکبیر
سارا مطیعی، دانشکده مهندسی کامپیوتر و فناوری اطلاعات دانشگاه صنعتی امیرکبیر

چکیده:

پیمایش وب برای جمع آوری صفحاتی به کار می رود که توسط موتور جستجو شاخص گذاری می شوند. موتور جستجو برای پاسخ دادن به پرس و جوی کاربران از این صفحات پیمایش شده و شاخص گذاری شده استفاده می کند. از آن جا که حجم صفحات وب بسیار بالا و همواره رو به افزایش است، موتورهای جستجو قادر به شاخص گذاری صفحات محدودی هستند. به همین جهت در سال های اخیر به جای پیمایش کل وب، الگوریتم های پیمایش موضوعی یا متمرکز مورد توجه محققین قرار گرفته است که در حین پیمایش 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 یادگیر سلولی ناهمگام برای پیمایش موضوعی 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 یادگیر سلولی ناهمگام، رفتار کاربران در مشاهده صفحات وب و پیوندهای بین صفحات، ساختار ارتباطی 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 ارتباطی به دست آمده از مرحله قبل، پیمایش موضوعی وب انجام می شود و صفحات مرتبط با یک موضوع خاص جمع آوری می گردد. نتایج ارزیابی پیمایشگر پیشنهادی حاکی از بهبود عملکرد آن (نرخ حاصل و فراخوان هدف) نسبت به پیمایشگر Best First و عدم وابستگی آن به انتخاب مجموعه آغازین دارد.