

تشخیص اجتماعات وب با استفاده از اتوماتای یادگیر توزیع شده و

پارتیشن بندی گراف

مجید تاران^۱، شهرزاد معتمدی مهر^۲، علی برادران هاشمی^۳، محمدرضا میبیدی^۴

چکیده

مجموعه ای از صفحات وب که درباره یک موضوع مشترک می باشند و توسط افراد یا سازمان های مختلف که علایق مشترک درباره آن موضوع خاص دارند ایجاد شده اند، اجتماع وب نامیده می شود. از آنجا که امروزه حجم وب از یک بیلیون صفحه گذشته است و همچنان در حال افزایش است، تشخیص اجتماعات وب روز به روز دشوارتر می شود. در این مقاله روشی ترکیبی مبتنی بر اتوماتای یادگیر توزیع شده و پارتیشن بندی گراف برای تشخیص اجتماعات وب پیشنهاد می گردد. روش پیشنهادی همان الگوریتم HITS می باشد که در آن علاوه بر ساختار پیوند بین صفحات، رفتار کاربر در مشاهده این صفحات نیز در نظر گرفته شده است. برای این منظور از اتوماتای یادگیر توزیع شده برای یادگیری امتیازات Hub و Authority صفحات وب استفاده می گردد. اجتماع وبی که به این روش به دست می آید وابسته به ساختار گرافی وب نمی باشد. الگوریتم پیشنهادی با استفاده از پیوندهای بین صفحات و رفتار کاربران وب، میزان شباهت و ارتباط بین صفحات را تعیین می کند. در این مقاله از پارتیشن بندی گراف وب برای بهبود کارایی استفاده شده است. به منظور ارزیابی، روش پیشنهادی پیاده سازی گردیده و نتایج آن با نتایج الگوریتم HITS و الگوریتمی دیگر مبتنی بر اتوماتای یادگیر توزیع شده مقایسه شده است. نتایج آزمایشها حاکی از کارایی روش پیشنهادی دارد.

کلمات کلیدی

اتوماتای یادگیر، اجتماع وب، داده های استفاده از وب

کنفرانس داده کاوی ایران

^۱ دانشکده برق و مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه آزاد قزوین، قزوین، ایران، m_taran@isc.iranet.net

^۲ دانشکده برق و مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه آزاد قزوین، قزوین، ایران، motamedi@tmu.ac.ir

^۳ دانشکده مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه صنعتی امیرکبیر، تهران، ایران، a_hashemi@aut.ac.ir

^۴ دانشکده مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه صنعتی امیرکبیر، تهران، ایران، mmeybodi@aut.ac.ir