

تشخیص اجتماعات وب با استفاده از اتوماتای یادگیر توزیع شده و پارسیشن بندی گراف

مجید تاران^۱، شهرزاد معتمدی مهر^۲، علی برادران هاشمی^۳، محمد رضا میبدی^۴

چکیده

مجموعه‌ای از صفحات وب که درباره یک موضوع مشترک می‌باشند و توسط افراد یا سازمان‌های مختلف که علاوه‌بر مشترک درباره آن موضوع خاص دارند ایجاد شده‌اند، اجتماع وеб نامیده می‌شود. از آنجا که امروزه حجم وеб از یک بیلیون صفحه گذشته است و همچنان در حال افزایش است، تشخیص اجتماعات وеб روز به روز دشوار‌تر می‌شود. در این مقاله روشی ترکیبی مبتنی بر اتوماتای یادگیر توزیع شده و پارسیشن بندی گراف برای تشخیص اجتماعات وеб پیشنهاد می‌گردد. روش پیشنهادی همان الگوریتم HITS می‌باشد 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 Authority و صفحات وеб استفاده می‌گردد. اجتماع وебی که به این روش به دست می‌آید وابسته به ساختار گرافی وеб نمی‌باشد. الگوریتم پیشنهادی با استفاده از پیوندهای بین صفحات و رفتار کاربران وеб، میزان شباهت و ارتباط بین صفحات را تعیین می‌کند. در این مقاله از پارسیشن بندی گراف و 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 اتوماتای یادگیر توزیع شده مقایسه شده است. نتایج آزمایشها حاکی از کارایی روش پیشنهادی دارد.

کلمات کلیدی

اتوماتای یادگیر، اجتماع وеб، داده‌های استفاده از وеб

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، motamed@tmut.ac.ir

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

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