

## عنوان مقاله:

ارائه مدلی برای بازیابی اطلاعات متنی با استفاده از اعداد فاصله ای

## محل انتشار:

دوفصلنامه فناوری اطلاعات و ارتباطات ایران، دوره 12، شماره 45 (سال: 1400)

تعداد صفحات اصل مقاله: 25

## نویسنده:

فرزاد قهرمانی - دانشجوی دکتری بخش مهندسی و علوم کامپیوتر و فناوری اطلاعات، دانشکده مهندسی برق و کامپیوتر، دانشگاه شیراز

## خلاصه مقاله:

با گسترش و توسعه وب و افزایش محتوای آنلاین، اهمیت سیستم های بازیابی اطلاعات که بتوانند با دقت بالاتری به نیازهای اطلاعاتی کاربران پاسخ دهند، بیشتر از پیش مشخص است. یک بخش مهم در طراحی هر سیستم بازیابی اطلاعات، انتخاب روشی مناسب برای مدل کردن آن سیستم است که در این راستا تعیین روش وزن دهی به لغات جهت بیان میزان اهمیت آنها در اسناد و پرس وجوها، نقش به سزائی دارد. روش های مختلفی در خصوص چگونگی وزن دهی به لغات ارائه شده که غالباً یک وزن عددی را تخصیص می دهند اما نمی توان با قطعیت گفت که بهترین روش وزن دهی کدام است. با توجه به ابهام و عدم قطعیتی که در این زمینه وجود دارد، در این مقاله مدلی ارائه شده که به جای استفاده از یک مقدار وزنی، با استفاده از وزن های بدست آمده از تعدادی روش وزن دهی پایه که به دقت انتخاب شده اند، برای هر لغت بازه ای از وزن ها را به عنوان یک وزن فاصله ای محاسبه می کند. در این مدل با انجام تجمیع مناسب، میزان ارتباط هر سند با پرس-وجوی ورودی نیز به صورت یک وزن فاصله ای تعیین شده و برحسب آنها می توان با استفاده از یکی از سه روش پیشنهادی، اسناد را رتبه بندی کرد. در آزمایش های انجام شده بر روی مجموعه داده های معتبر Cranfield و Medline، اثرات نرمال سازی طول بردار وزن های پایه، استفاده از مولفه های مختلف در فاکتور فرکانس لغت و فاکتور فرکانس مجموعه مورد مطالعه و بحث قرار گرفته است و مشخص شد که انتخاب مجموعه ای مناسب از روش های وزن دهی پایه برای اعمال روش پیشنهادی، به همراه استفاده از روش رتبه بندی مناسب، تاثیر به سزائی در بهبود بازدهی سیستم خواهد داشت. با انتخاب های مناسب، برای دو مجموعه داده مذکور به ترتیب MAP با مقادیر ۰.۴۳۳۲۳ و ۰.۵۴۵۸۰ بدست آمد. این نتایج نشان داد که روش پیشنهادی نه تنها باعث بهبود نسبت به هر یک از روش های وزن دهی پایه می شود، بلکه در مقایسه با چند روش وزن دهی پیچیده اخیر نیز بهتر عمل می کند.

## کلمات کلیدی:

بازیابی اطلاعات متنی، رتبه بندی اسناد، وزن دهی لغات، اعداد فاصله ای، وزن فاصله ای

## لینک ثابت مقاله در پایگاه سیویلیکا:

<https://civilica.com/doc/1858937>

