

عنوان مقاله:

روشی جهت حذف وندهای کلمات فارسی با استفاده از برچسب دستوری نوع کلمات

محل انتشار:

یازدهمین کنفرانس سراسری سیستم های هوشمند (سال: 1391)

تعداد صفحات اصل مقاله: 6

نویسندگان:

رضا مصلح فسایی - دانشکده آموزش های الکترونیکی، دانشگاه شیراز

محمد هادی صدرالدینی - بخش مهندسی و علوم کامپیوتر و فناوری اطلاعات، دانشکده برق و کامپیوتر، دانشگاه شیراز

سید مصطفی فخر احمد - بخش مهندسی و علوم کامپیوتر و فناوری اطلاعات، دانشکده برق و کامپیوتر، دانشگاه شیراز

خلاصه مقاله:

هدف از این مقاله، ارائه روشی جهت حذف «وندهای» کلمات فارسی می باشد. در زبان فارسی «وندهای» مختلفی وجود دارند که به قبل یا بعد از کلمات افزوده می شوند. «وندها» باعث ایجاد هماهنگی کلمات، در جمله می شوند اما در تغییر ماهیت کلمات تأثیری ندارند. در برخی از سیستم های پردازش متن فارسی از جمله سیستم های طبقه بندی کننده متون، به الگوریتم نیاز است که از طریق آنها، وندهای فارسی را حذف می کنیم تا متون، ساده تر شوند. در این مقاله ابتدا به کلمات فاقد وند موجود در زبان فارسی، برچسب دستوری می زنیم. برچسب دستوری، نوع کلمه را مشخص می کند. نوع کلمه می تواند اسم، بن فعل، وندها و ... باشد. سپس با شناسایی طولانی ترین توالی از وندهایی که طبق دستور زبان فارسی به انواع مختلف کلمات متصل می شوند، اقدام به حذف زیر توالی وندهای متصل به کلمات می کنیم. جهت تولید زیر توالی وندهای کلمات فارسی، آنها را با استفاده از یک روال بازگشتی پردازش کرده و توالی نوع زیر کلماتشان را استخراج می کنیم. حذف توالی های وندهای مجاز، با توجه نوع کلمه یا بعد از توالی وندها انجام می شود. این روش چون حذف وندها را با آگاهی از نوع کلمات انجام می دهد، دارای دقت 99% می باشد

کلمات کلیدی:

برچسب دستوری نوع کلمه، توالی وندها، نوع کلمات

لینک ثابت مقاله در پایگاه سیویلیکا:

<https://civilica.com/doc/214650>

