

عنوان مقاله:

برچسب زنی آماری کلمات فارسی با استفاده از وندها

محل انتشار:

اولین کنفرانس ملی محاسبات نرم و فن آوری اطلاعات (سال: 1389)

تعداد صفحات اصل مقاله: 5

نویسندگان:

علی اصغر بهمنش - گروه کامپیوتر - دانشگاه بوعلی سینا

امیر سالاریور - گروه کامپیوتر - دانشگاه بوعلی سینا

عبدالحمید پیله ور - گروه کامپیوتر - دانشگاه بوعلی سینا همدان

خلاصه مقاله:

یکی از اقدامات اساسی در حوزه پردازش زبانهای طبیعی برچسب زنی کلمات می باشد. برچسب ها در حوزه های پردازش گفتار (تبدیل متن به گفتار و برعکس)، ماشین مترجم [ریشه یاب ها و دیگر کاربردهای پردازش زبان استفاده می شود. در این مقاله سعی شده است تا کلمات به وسیله تکنیک های آماری روی پیکره فارسی بیجین خان به درستی برچسب گذاری شوند. روش های آماری یکی از ابتدایی ترین روش برچسب زنی است که علاوه بر سادگی، دقت قابل قبولی نیز دارد. برای اینکه برچسب صحیحی به کلمات داده شود. لازم است کلمات از نظر ساختاری مورد بررسی و پردازش قرار بگیرند. به همین دلیل قبل از برچسب زنی کلمات، ساختار کلمه از لحاظ پیشوندی و پسوندی پیش پردازش شده اند و برچسب با توجه به وندهای موجود در کلمه انتخاب می شود. بنابراین دقت برچسب زنی افزایش می یابد. روش ارائه شده در این مقاله 96.07% صحت را برای برچسب زنی روی پیکره نشان می دهد که نسبت به روش های مشابه دقتی بهتری دارد.

کلمات کلیدی:

پردازش زبان طبیعی، برچسب زنی کلمات، پردازش زبان فارسی، تخمین بیشینه احتمال

لینک ثابت مقاله در پایگاه سیویلیکا:

<https://civilica.com/doc/132778>

