

عنوان مقاله:

بررسی تاثیر دقت برچسب اجزای کلام در کارایی سامانه شناسایی موجودیت های نامدار برای زبان فارسی

محل انتشار:

اولین همایش جویشگر بومی (سال: 1394)

تعداد صفحات اصل مقاله: 4

نویسندگان:

شادی حسین نژاد - گروه پردازش صوت و زبان طبیعی، پژوهشگاه توسعه فناوری های پیشرفته خواجه نصیرالدین طوسی، تهران،

یاسر شکفته - گروه پردازش صوت و زبان طبیعی، پژوهشگاه توسعه فناوریهای پیشرفته خواجه نصیرالدین طوسی، تهران،

طاهره امامی آزادی - گروه پردازش صوت و زبان طبیعی، پژوهشگاه توسعه فناوریهای پیشرفته خواجه نصیرالدین طوسی، تهران

خلاصه مقاله:

تشخیص موجودیت های نامدار یکی از مسایل پردازش زبان طبیعی است که هدف آن شناسایی موجودیت های نامدار موجود در یک متن و دسته بندی آنها در دسته های از پیش تعیین شده است. یکی از روش های تشخیص موجودیت نامدار، استفاده از پیکره برچسبگذاری شده و اعمال روش های یادگیری ماشین است. در این مقاله نحوه تولید یک سامانه تشخیص موجودیت نامدار مبتنی بر پیکره با استفاده از روش میدان های تصادفی شرطی شرح داده میشود. همچنین تاثیر دقت برچسب اجزا کلام (به عنوان یک ویژگی مورد استفاده) بر دقت سامانه موجودیت نامدار بررسی میگردد. سامانه ذکر شده با استفاده از پیکره فارسی اعلام که حاوی حدود 250 هزار توکن است، تولید شده است. آموزش این سامانه با استفاده از برچسب های اجزای کلام دقیق برگرفته از پیکره متنی فارسی منتج به f-85 measure درصد شده است. این در حالی است که استفاده از برچسب های تولید شده از یک ابزار برچسب گذاری اجزای کلام (برچسب گذاری غیر دقیق) در آموزش این سامانه منجر به دستیابی به f-measure 84/9 درصد شده است.

کلمات کلیدی:

تشخیص موجودیت های نامدار، پیکره موجودیت های نامدار، پردازش زبان طبیعی، یادگیری ماشین، برچسب اجزای کلام

لینک ثابت مقاله در پایگاه سیویلیکا:

<https://civilica.com/doc/754998>

