

عنوان مقاله:

سازوکاری برای بهینه سازی استخراج ویژگی برای تشخیص هممرجعی ضمیر و اسم اشاره در زبان فارسی

محل انتشار:

پنجمین کنفرانس مهندسی دانش بنیان و نوآوری (سال: 1397)

تعداد صفحات اصل مقاله: 8

نویسندگان:

لادن مداح - دانشکده مهندسی کامپیوتر، دانشگاه علم و صنعت استان تهران تهران، ایران

بهروز مینایی - دانشکده مهندسی کامپیوتر، دانشگاه علم و صنعت استان تهران تهران، ایران

خلاصه مقاله:

تشخیص هم مرجعی یکی از پایه های مهم در پردازش زبان طبیعی می باشد. کاربردهای مهمی در حوزه هایی مانند پاسخ گویی به سوال، ترجمه ماشین، خلاصه سازی اتوماتیک و استخراج موجودیت نامدار دارد. وظیفه ی تشخیص هم مرجعی حل و فصل عبارت های اسمی و ضمائر در متن است که به موجودیت یکسان ارجاع می دهد. روش های تشخیص هم مرجع را می توان به دو دسته ی روش های زبان شناسی و روش های یادگیری ماشین تقسیم نمود. روش های زبان شناسی بیشتر به اطلاعات زبان شناسی نیاز دارند، البته مشکلی که این روش ها دارند این است که احتمال خطا در آنها بیشتر است همچنین اجرای این روش ها زمان بر می باشند، در حالی که روش های یادگیری ماشین کمتر به اطلاعات زبان شناسی نیاز دارند. در این مقاله سیستمی ارائه می شود که اسم های هم مرجعی موجود در متن را شناسایی کرده و با در نظر گرفتن ویژگی های مشخص شده نمونه های مثبت و منفی را از پیکره استخراج می کند. از جمله نوآوری هایی که انجام شد اضافه کردن ویژگی هایی به سیستم می باشد که از آن ها برای عمل تشخیص مرجع ضمیر استفاده می شود. این ویژگی ها از نقش نحوی عبارت اسمی کاندید ضمیر استفاده کرده است. با تغییر دادن این ویژگی ها در سیستم در بهترین حالت میتوان به دقت 59.4 درصد دست پیدا کردیم.

کلمات کلیدی:

پردازش زبان طبیعی، پیکره زبان فارسی، تشخیص مرجع ضمیر، یادگیری ماشین

لینک ثابت مقاله در پایگاه سیویلیکا:

<https://civilica.com/doc/988924>

