

عنوان مقاله:

ارائه یک روش جدید برای تولید سوپرسن س در شبکه واژگانی وردنت

محل انتشار:

پنجمین کنفرانس بین المللی پژوهش های کاربردی در کامپیوتر، برق و فناوری اطلاعات (سال: 1399)

تعداد صفحات اصل مقاله: 15

نویسندگان:

مهرداد محمدیان - دانشکده مهندسی کامپیوتر، دانشگاه علم و صنعت ایران

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

خلاصه مقاله:

وردنت به عنوان یکی از بزرگترین پایگاه داده های واژگانی، در بسیاری از پژوهش های مربوط به پردازش زبان طبیعی مورد استفاده قرار گرفته است. این شبکه واژگانی بزرگ و پیچیده علاوه بر کاربردهایی که تا به امروز داشته است از نگاهبرخی پژوهشگران دارای نقاط ضعفی نیز می باشد. این گروه از پژوهشگران بر این باور هستند که شبکه واژگانی وردنت به صورت ریزدانه است که همین امر می تواند سبب کاهش کارایی در برخی از کارهای مرتبط با پردازش زبان طبیعی از جمله بازیابی اطلاعات، ترجمه ماشینی، گسترش پرس وجو و ابهام زدها شود. از سویی، وردنت به هر یک از هم نشیم های موجود در درون خود و در زمان تولید یک برچسب اختصاص داده است که این عمل تماماً توسط زبان شناسان صورت پذیرفته است. مجموع برچسب های اختصاص داده شده به هم نشیم ها عدد 45 را نشان می دهد که بیانگر این موضوع است که می توان به جای 117 هزار هم نشیم، 45 هم نشیم بزرگتر به نام سوپرسنس را در بسیاری از کارهای علمی در نظر گرفت. ما در این پژوهش و در یک رویکرد جدید اقدام به تولید 45 سوپرسنس جدید کردیم. برای مقایسه کیفیت سوپرسنس های ایجاد شده از یک ابهام زدا به نام UKB که بالاترین دقت در میان ابهام زدهای مبتنی بر متن را در اختیار دارد، بهره گرفتیم. نتایج آزمایش ها نشان می دهند که سنجه-اف حاصل شده از سوپرسنس های ارائه شده توسط ما بهتر از سوپرسنس های ایجاد شده توسط انسان عمل کرده است. همچنین، با توجه به کاهش ابعاد گراف وردنت سرعت عملکرد ابهام زدا بیش از پانصد برابر سریع تر از زمانی است که از وردنت استاندارد با 117 هزار هم نشیم استفاده می کند و این در حالی است که دقت نهایی تنها در حدود یک درصد کاهش می یابد.

کلمات کلیدی:

وردنت، شباهت، پردازش زبان طبیعی، ابهام زدا

لینک ثابت مقاله در پایگاه سیویلیکا:

<https://civilica.com/doc/1123549>

