

عنوان مقاله:

ریشه یابی کلمات با استفاده از شبکه ی بیزین ساده و کاربرد آن به همراه استفاده از یک الگوریتم جدید در تعیین میزان مشابهت متون

محل انتشار:

یازدهمین کنفرانس سراسری سیستم های هوشمند (سال: 1391)

تعداد صفحات اصل مقاله: 8

نویسندگان:

فاطمه کاظمی ونهری - دانشگاه آزاد اسلامی واحد علوم و تحقیقات تهران

علی احمدی - دانشگاه صنعتی خواجه نصیرالدین طوسی

مهران محسن زاده - دانشگاه آزاد اسلامی واحد علوم و تحقیقات تهران

خلاصه مقاله:

امروزه از ریشه یاب ها استفاده های بسیاری در مسائل مربوط به حوزه زبان طبیعی، داده کاوی و بازیابی اطلاعات میشود. از جمله مسائل مورد بحث در حوزه ی زبان طبیعی، مسئله ی تشخیص میزان مشابهت متون است. قبل از تشخیص میزان مشابهت متون، عملیاتی تحت عنوان پیش پردازش بر روی متون انجام می گیرد. پیش پردازش ها شامل انجام فعالیت هایی بر روی متن است که باعث می شوند نتایج الگوریتم پیدا کردن تشابه، بهتر گردد. پیش پردازشها به طور کلی شامل مراحل حذف علائم نقطه گذاری، جداسازی کلمات، حذف اعداد، حذف کلمات پرتکرار و ریشه یابی است. در این مقاله پس از انجام چهار مرحله ی ابتدایی پیش پردازش، با ارائه ی یک الگوریتم جدید مبتنی بر شبکه ی بیزین ساده، به انجام یک مرحله ی اضافی دیگر تحت عنوان دسته بندی کلمات می پردازیم و از نتایج دسته بندی در مرحله ی ریشه یابی استفاده می کنیم. سپس به بررسی کاربرد آن الگوریتم به همراه استفاده از یک الگوریتم جدید بر مبنای تئوری احتمال کل، و فاصله یابی در فضای برداری، به تشخیص میزان مشابهت متون می پردازیم. بررسی های ما نشان می دهد که الگوریتم های ارائه شده توانسته اند نتایج بسیار خوبی را در زمینه ی ریشه یابی و تعیین میزان مشابهت متون بدست آورند

کلمات کلیدی:

دسته بندی کلمات، ریشه یابی، شینگل، فضای برداری، میزان مشابهت

لینک ثابت مقاله در پایگاه سیویلیکا:

<https://civilica.com/doc/214689>

