

عنوان مقاله:

بازشناسی زیرکلمات تایپی فارسی با یک روش سلسله مراتبی

محل انتشار:

سومین کنگره بین المللی کامپیوتر، برق و مخابرات (سال: 1395)

تعداد صفحات اصل مقاله: 23

نویسندگان:

اسماعیل میری - دانشجوی دکتری دانشکده مهندسی برق و کامپیوتر، دانشگاه بیرجند

سیدمحمد رضوی - دانشیار دانشکده مهندسی برق و کامپیوتر، دانشگاه بیرجند

ناصر مهرشاد - دانشیار دانشکده مهندسی برق و کامپیوتر، دانشگاه بیرجند

خلاصه مقاله:

در این مقاله روشی ساده برای بازشناسی زیرکلمات تایپی فارسی ارائه می شود. ابتدا با استفاده از چندویژگی ساده و با توجه به موقعیت نقاط و علائم زیرکلمه ورودی، فضای جستجو محدود می شود. سپس با استفاده از ویژگی های مکان مشخصه دو زیرکلمه که کمترین فاصله را با زیرکلمه ورودی دارند انتخاب می شوند. دو زیرکلمه نیز با استفاده از ویژگی های ناحیه بندی و معیار کمترین فاصله انتخاب می شوند که ممکن است بین این دو و دو زیرکلمه قبلی یک یا دو زیرکلمه مشترک نیز وجود داشته باشد. برای انتخاب زیرکلمه نهایی از بین این حداکثر چهار زیرکلمه، از ویژگی های تعداد تقاطع عمودی متن با زمینه استفاده شده است. این روش برای قلم لوتوس آزمون شده و نرخ بازشناسی 98.18% برای داده های آزمون بدست آمده است.

کلمات کلیدی:

بازشناسی زیرکلمات فارسی، کاهش فضای جستجو، ویژگی های ساده، ویژگی های مکان مشخصه، ویژگی های ناحیه بندی، ویژگی های تقاطع عمودی

لینک ثابت مقاله در پایگاه سیویلیکا:

<https://civilica.com/doc/576224>

