

عنوان مقاله:

بهبود بازشناسی متن فارسی با استفاده از اطلاعات در سطح کلمات

محل انتشار:

سومین کنفرانس بین المللی بازشناسی الگو و تحلیل تصویر ایران (سال: 1396)

تعداد صفحات اصل مقاله: 7

نویسندگان:

بی تا نامور - دانشجو مقطع کارشناسی ارشد دانشگاه الزهرا (س)

رضا عزمی - دانشیار گروه مهندسی کامپیوتر دانشگاه الزهرا (س)

خلاصه مقاله:

در این مقاله روشی برای بهبود بازشناسی تصاویر متون چاپی فارسی با استفاده از اطلاعات آماری زبان در سطح کلمات ارایه شده است. در این روش ابتدا تصویر براساس روش مبتنی بر شکل کلی زیرکلمات بازشناسی می شود و برای هر زیرکلمه، زیرکلمات محتمل دیگر از خروجی سیستم OCR دریافت می گردند، سپس تصحیح خطا با استفاده از اطلاعات مدل زبانی بایگرم در سطح کلمه انجام می گردد. در این روش برای به دست آوردن کلمات محتمل از مدل مخفی مارکوف استفاده شده است. برای آزمایش روش های پیشنهادی، 19 سند با چهار فونت B Nazanin, B Mitra, B Lotus, و B Yagut با اندازه های 14، 16، 18، 20 فراهم گردید. بازشناسی در سطح زیر کلمه با دقت حدود 88 درصد و اصلاح خطا در سطح کلمات دقت را به حدود 92 درصد افزایش داده است.

کلمات کلیدی:

OCR، پردازش زبان طبیعی، مدل مخفی مارکوف، مدل بایگرم، متن چاپی

لینک ثابت مقاله در پایگاه سیویلیکا:

<https://civilica.com/doc/638509>

