

عنوان مقاله:

یک رویکرد مبتنی بر مدل برای شناسایی و اصلاح نویسه های فارسی و عرب مستقل از متن آفلاین

محل انتشار:

دومین کنگره سراسری فناوریهای نوین ایران با هدف دستیابی به توسعه پایدار (سال: 1394)

تعداد صفحات اصل مقاله: 23

نویسندگان:

سید محمد سلیمانی - مدرس (مربی)، گروه فناوری اطلاعات، مرکز آموزش علمی کاربردی شهرداری بیارجمند

هادی حسینی - دانشجو، گروه فناوری اطلاعات، مرکز آموزش علمی کاربردی شهرداری بیارجمند

خلاصه مقاله:

مدل های نسل حرکت سریع در مقالات به عنوان یک ابزار موثر برای درک رفتار دست خط توصیف شده است. زمینه های کاربرد گوناگون هستند که شامل توصیف دست خط بازسازی و اخیراً OCR، در این مقاله یک رویکرد مبتنی بر نویسه (یکی از حروف الفبا) را برای شناسایی اصلاح نویسه فارسی و عربی آفلاین پیشنهاد می کنیم. بلکه بیش از استخراج نویسه های خودش را براساس مدل بتا-بیضی ترکیب می کند اصل آن متکی به استقلال از کتاب کد نویسه از همه فرایند تمرینی و استفاده از یک مدل به جای آن است. یک کتاب کد کامل و چهار کتاب کد جزئی درست شده و تست شده اند، با استفاده از انتخاب مشخصه، کتاب های کد خام با توجه به FDR، و رابطه ی مقطعی و معیار نمونه گیری تصادفی، سائز آن ها کاهش می یابد. یک مجموع 60 بردار مشخصه با استفاده از تطبیق الگو استخراج شده اند و با 411 نویسه جداگانه از دیتابیس FN/ENIT1 ارزیابی شدند. نتایج ارائه شده در این مطالعه نمایش وسیع قابلیت تعمیم دهی خوب از کتاب های کد ترکیبی را نشان می دهد. ما یک نرخ 90/02% در نخست و سپس نرخ 96/35% برای شناسایی نویسنده و یک $EER=2/1$ برای اصلاح نویسه بدست آوردیم رویکرد ما، خصوصیات بهتری نسبت به بیشتر تکنیک های بررسی شده برحسب سائز نوشته جات و نرخ شناسایی، نشان داد. تا آنجا که می دانیم این مطالعه اولین مطالعه برای درونیابی مفهوم کتاب های کد ترکیبی مبتنی بر مدل در شناسایی و اصلاح نویسه های فارسی و عربی است.

کلمات کلیدی:

شناسایی و اصلاح، متن آفلاین، لغات فارسی، الفبای، بیومتریک، رویکرد شناسایی، دیتابیس، ماتریس، مدل بتا، شکل بیضی

لینک ثابت مقاله در پایگاه سیویلیکا:

<https://civilica.com/doc/399667>

