

عنوان مقاله:

تقسیم بندی و تشخیص زیرکلمات اسناد قدیمی دست نویس فارسی

محل انتشار:

فصلنامه دستاوردهای نوین در برق، کامپیوتر و فناوری، دوره 2، شماره 5 (سال: 1401)

تعداد صفحات اصل مقاله: 17

نویسندگان:

سمیه صبوری - دانشجوی دکترا، دانشگاه آزاد اسلامی واحد بیرجند

حمیدرضا غفاری - استادیار دانشگاه آزاد اسلامی واحد فردوس

خلاصه مقاله:

اسناد تاریخی همواره مورد توجه مورخان و زبان شناسان است. اسناد مهم معمولاً توسط روش های تقسیم بندی و شناسایی به صورت دیجیتالی تبدیل می شود. دیجیتالی کردن اسناد برای تحقیق بر روی این اسناد و حفاظت از آنها اهمیت فراوانی دارد. این مقاله یک چهارچوب تقسیم بندی و تشخیص کلی برای تصاویر اسناد تاریخی فارسی پیشنهاد شده است. ابتدا با حذف نویزها، رفع کجی، حذف مهرها و ... پیش پردازش روی اسناد انجام شده و تصویر سند به یک تصویر دو سطحی تبدیل می شود. در مرحله دوم یک روش تقسیم بندی متن سند به خطوط پیشنهاد شده است. در مرحله سوم یک روش تقسیم بندی خطوط به زیرکلمات رسم الخط فارسی را ارائه کرده و زیرکلمات این اسناد را استخراج نموده سپس از شبکه های عمیق برای آموزش زیرکلمات پرتکرار و تشخیص آنها استفاده شده و نتایج بر مبنای معیارهای مختلف گزارش شده است.

کلمات کلیدی:

اسناد دست نویس فارسی، تقسیم بندی سند، تقسیم بندی خطوط، تقسیم بندی زیرکلمات، شبکه عصبی

لینک ثابت مقاله در پایگاه سیویلیکا:

<https://civilica.com/doc/1746495>

