

عنوان مقاله:

مروری بر روش های جستجوی کلمه در اسناد تصویری

محل انتشار:

نخستین همایش ملی دستاوردهای نوین در مهندسی برق، مهندسی کامپیوتر و مهندسی پزشکی (سال: 1402)

تعداد صفحات اصل مقاله: 15

نویسندگان:

مسلم محمدی - استادیار گروه مهندسی کامپیوتر و فناوری اطلاعات، دانشکده فنی و مهندسی، دانشگاه پیام نور تهران

مجید ایرانپورمبارکه - استادیار گروه مهندسی کامپیوتر و فناوری اطلاعات، دانشکده فنی و مهندسی، دانشگاه پیام نور تهران

خلاصه مقاله:

کتابخانه های سرتاسر دنیا حاوی اطلاعات باارزشی هستند که با پیشرفت فناوری دیجیتال و جهت حفظ این منابع باارزش، به صورت اسکن شده (تصویر) درآمده اند. جستجو روی این اسناد تصویری کار بسیار زمان بری است. یک راه حل، استفاده از بازشناسی کاراکتر نوری (OCR) برای تبدیل اسناد کاغذی اسکن شده به اسناد متنی و استفاده از موتورهای بازیابی متن است. فناوری های بازشناسی کاراکتر نوری موجود در مورد فونت های استاندارد با پس زمینه تمیز خوب کار می کند، اما زمانی که کیفیت اسناد چالی افت می کند و همچنین در مورد اسناد دست-نویس نتایج بسیار ضعیفی را نشان می دهد. رهیافت جایگزین برای اندیس گذاری متون چاپی قدیمی و دست نویس جستجوی کلمات است. در این مقاله، سعی شده است تا مروری کامل بر روش های جستجوی کلمه در اسناد تصویری انجام گیرد. دسته بندی های مختلف از دیدگاه های متخصصین این حوزه و همچنین مزایا و معایب تکنیک های ارائه شده تا به امروز، جهت ایجاد بینش کامل در این زمینه، خصوصا جهت کاربرد در اسناد فارسی، آورده شده است.

کلمات کلیدی:

جستجوی کلمه تصویری، بازیابی کلمه تصویری، اندیسگذاری اسناد تصویری، بازیابی مبتنی بر محتوای تصویر

لینک ثابت مقاله در پایگاه سیویلیکا:

<https://civilica.com/doc/1852318>

