

## عنوان مقاله:

آنالیز و بازیابی مستندات فارسی با استفاده از قطعه بندی صفحه مستندات

## محل انتشار:

اولین کنفرانس داده کاوی ایران (سال: 1386)

تعداد صفحات اصل مقاله: 15

## نویسندگان:

حمیدرضا مقسمی - عضو هیات علمی دانشگاه آزاد اسلامی واحد تهران سما گروه کامپیوتر

علی برومندیا - استادیار دانشگاه آزاد اسلامی واحد تهران جنوب گروه کامپیوتر

آذین پشتیار - کارشناسی کامپیوتر گرایش سخت افزار دانشگاه آزاد اسلامی واحد تهران جنوب

## خلاصه مقاله:

آنالیز قطعه بندی تصویر مستند، پردازش شناسایی ساختارهای قطعه بندی توسط آنالیز تصویر مستند است. با تحلیل محتویات متون قطعه بندی شده و به کمک الگوریتم های بازیابی می توان متون قطعه بندی شده را به بخش های مختلف تفکیک نمود و از این طریق تکنیک سودمندی را جهت جستجوی هر چه سریع تر و دقیق تر بخش های گوناگون مستندات فارسی بر مبنای تصاویر بازیابی و پردازش شده، در پیش گرفت. این مقاله برای قطعه بندی صفحه مستندات فارسی، روش ترکیبی در رزولوشن پائین و در رزولوشن بالا، را ارائه می نماید. در آنالیز چند رزولوشنی، سطوح رزولوشن پایین در ساختار تصویر هرمی برای آنالیز کل تصویر استفاده می شود و تصویر باینری مستند به مجموعه ای از نواحی مجزا قطعه بندی می گردد. در قطعه بندی صفحه در روش رزولوشن بالا، همه نواحی در تصویر مستند به نواحی متنی، عکس، ترسیمی و جداول قطعه بندی می شوند. همچنین می توان با این راهکار با توجه به کاربرد مورد نظر خود، کلمات و اشکال و نمودارهای خاصی را در مستندات فارسی جستجو و بازیابی نمود. روش پیشنهاد شده روی مجموعه ای از صفحات مستندات فارسی مورد آزمایش قرار گرفته است. نتایج این آزمایشات توانائی و برتری روش ما را در مقایسه با روش های قبلی نشان می دهد.

## کلمات کلیدی:

پردازش تصویر مستندات، قطعه بندی صفحات مستندات فارسی، قطعه بندی مبتنی بر شناسائی، مؤلفه های همبند

## لینک ثابت مقاله در پایگاه سیویلیکا:

<https://civilica.com/doc/33105>

