

## عنوان مقاله:

استخراج موضوع و زیر موضوع در سیستم هوشمند موضوع بندی متن متراکم فارسی (نصیر)

## محل انتشار:

اولین کنفرانس بین المللی فناوری اطلاعات و دانش (سال: 1382)

تعداد صفحات اصل مقاله: 8

## نویسندگان:

ناصر قاسم آقایی - گروه کامپیوتر دانشگاه اصفهان

قاسم صرافان

## خلاصه مقاله:

در این مقاله برای اولین بار پرونده های متنی از لحاظ تراکم موضوعی، بررسی گردیده و به دو دسته: متنهای متراکم و نامتراکم تفکیک شده اند. متن متراکم متنی است، که شامل دامنه وسیعی از موضوعات باشد. یا به عبارت دیگر میزان چگالی موضوعی (تعداد موضوعات تقسیم بر طول پرونده) آنها بالا باشد (مانند قرآن کریم، نهج البلاغه، دیوانهای شعر، آرشیو مجلات و دایره المعارف ها). در این تحقیق از روابط وابستگی، مسیرها و پایگاه ه داده ای از سه تایی ها و همچنین روشهای آماری پردازش متن، در جهت استخراج کلمات و ترکیبات برجسته از متنهای متراکم، و قرار دادن آنها در یک سلسله مراتب خوشه ای، استفاده می شود. همچنین در این راستا فرآیندی برای یافتن مرجع ضمیرها- که خاص متنهای متراکم است- تشریح خواهد شد. نتایج بدست آمده از نمونه پیاده سازی شده، نشانگر بهبود کیفیت جستجو و اندیس گذاری در متنهای متراکم زبان فارسی است.

## کلمات کلیدی:

متن متراکم فارسی، موضوع بندی، روابط وابستگی، ترکیبات برجسته، مرجع یابی ضمیر

## لینک ثابت مقاله در پایگاه سیویلیکا:

<https://civilica.com/doc/72121>

