

## عنوان مقاله:

استخراج کلیدواژگان پایان نامه فارسی با استفاده از ویژگی آماری و دسته بند بیز

## محل انتشار:

دوماهنامه جستارهای زبانی، دوره 12، شماره 6 (سال: 1400)

تعداد صفحات اصل مقاله: 29

## نویسندگان:

بهزاد حجازی - *M.Sc Student of Artificial Intelligence, Islamic Azad University North Tehran Branch, Tehran, Iran*

جلال الدین نصیری - *Department of Computational Linguistics, Information Science Research Department, Iranian - Research Institute for Information Science and Technology (IRANDOC), Tehran, Iran*

## خلاصه مقاله:

هدف اصلی استخراج کلمات کلیدی انتخاب مجموعه ای از لغات در متن است که می تواند موضوع اصلی متن را بازگو کند. استخراج کلیدواژگان در بازیابی اطلاعات، سیستم های پیشنهاددهنده متنی و دسته بندی متون، نقش مهم را ایفا می کند. در زبان فارسی باتوجه به پیچیدگی ذاتی زبان فارسی استخراج کلیدواژگان به مراتب دشوارتر شده است. در این پژوهش سعی شده است با رویکرد نوین ترکیبی آماری و یادگیری ماشین به استخراج کلیدواژگان پرداخته شود. ابتدا باتوجه به ساختار زبان فارسی پیش پردازش های لازم برای حذف کلمات و علائم نگارشی صورت می گیرد. سپس با استفاده از سه نوع ویژگی آماری و دسته بند بیز سیستم به صورت خودکار الگوی کلمات کلیدی با کلمات عادی را آموزش می بیند. همچنین پس پردازش کارا برای کم کردن کلمات مثبت کاذب در چارچوب پیشنهادی طراحی شده است. گفتنی است که مدل ساخته شده قادر به شناسایی تعداد حداکثر ۲۰ کلیدواژه در هر پایان نامه است و این کلمات با کلیدواژگان نوشته شده در هر متن مقایسه و ارزیابی می شوند. نتایج ارزیابی های متنوع نشان می دهد روش پیشنهادی با دقت مناسبی توانسته است کلمات کلیدی نوشتارهای فارسی علمی (پایان نامه و رساله) را استخراج کند.

## کلمات کلیدی:

,Extraction, Bayesian Classification, statistical features, preprocessing, post-processing  
استخراج کلیدواژگان، دسته بند بیز، ویژگی های آماری، پیش پردازش، پس پردازش

## لینک ثابت مقاله در پایگاه سیویلیکا:

<https://civilica.com/doc/1433522>

