

ارائه خلاصه ساز ترکیبی با هدف افزایش کیفیت و بهبود زمان

سجاد جهانگیری زرکانی، دکتر مهدی جوانمرد

۱- دانشجوی کارشناسی ارشد، دانشگاه پیام نور

۲- استادیار، دانشگاه پیام نور



esjay_5@yahoo.com

سجاد جهانگیری زرکانی

خلاصه

استخراج مطالب اصلی از یک متن این امکان را به انسان می دهد تا بتواند خلاصه ی متن را در حجمی بسیار کمتر از حجم اصلی مطالعه کند. این خلاصه باید دارای کیفیتی قابل پذیرش باشد. خلاصه سازی های موجود با دو روش یا ترکیبی از این دو کار می کنند. روش های نحوی و دیگری روش های معنایی. در روش معنایی، باید همه کلمات متن بررسی شوند و یا حتی همه کلمات با مجموعه کلمات دیگری مقایسه شوند. روش های معنایی زمان بر می باشد اما کیفیت خلاصه ساز را بهبود می بخشد. پارامتر زمان نیز فاکتور مهمی در تولید خلاصه می باشد. در این مقاله با استفاده از تکنیک های چند نخه الگوریتم ترکیبی ارایه شده است که هم از پارامترهای نحوی و هم پارامترهای معنایی استفاده می کند. نتایج نشان می دهد که علاوه بر افزایش کیفیت، زمان تولید خلاصه نیز حدوداً دارای تسریع ۴.۶ برابر می باشد

کلمات کلیدی: خلاصه سازی، خلاصه سازی کارآ، پارامترهای نحوی، پارامترهای معنایی

۱. مقدمه

خلاصه سازی متن فرایند شناسایی و استخراج برجسته ترین اجزای یک سند یا مجموعه ای از اسناد و گردآوری آنها در حجمی کمتر نسبت به متن اصلی می باشد که این اجزاء بیشترین تطابق و تعلق به موضوع و مفهوم مطرح شده در سند را داشته باشند [۱]. با توجه به مطالب فوق وجود سیستمی خودکار که توانایی خلاصه سازی متن را داشته باشد و انسان را در تجزیه و تحلیل این اسناد یاری کند بسیار مفید می باشد [۲][۳]. به طور کلی روش های خلاصه سازی را می توان به دو دسته ی extractive و abstraction تقسیم کرد. در روشهای extractive ویژگی هایی را برای هر قسمت از متن مشخص می کنند و براساس آن ویژگی ها و خصوصیات، میزان درجه ی اهمیت هر جزء از متن مشخص شده و در نهایت بهترین آن ها به عنوان اجزاء خلاصه نهایی انتخاب می شوند. در مقابل در روش های Abstraction معمولاً از تکنیک های مربوط به پردازش زبان طبیعی استفاده می شود و رسیدن به آن معمولاً به خاطر دشواری های مربوط به پردازش زبان طبیعی به طور کامل امکان پذیر نمی باشد. روش هایی که در این دسته انجام می شوند معمولاً ترکیبی از دو روش می باشند [۴]. بدلیل عدم وجود قطعی در خلاصه سازی متن و برای برخورد با عدم قطعیت از منطق فازی [۵] استفاده می شود. در واقع از منطق فازی برای اندازه گیری درجه ی اهمیت و میزان ارتباط و همبستگی و همچنین مشخص کردن جملات مهم برای ایجاد