

عنوان مقاله:

کاوشی بر رفتار الگوریتم های کلاسه بندی مبتنی بر درخت تصمیم گیری و الگوریتم نزدیکترین K همسایه در متون فارسی

محل انتشار:

همایش مهندسی کامپیوتر و توسعه پایدار با محوریت شبکه های کامپیوتری، مدلسازی و امنیت سیستم ها (سال: 1392)

تعداد صفحات اصل مقاله: 6

نویسندگان:

ندا صادقیان - دانشگاه آزاد اسلامی واحد علوم و تحقیقات گروه کامپیوتر، یزد، ایران

محمدجواد کارگر - استادیار دانشگاه آزاد اسلامی واحد میبد، گروه کامپیوتر، میبد، ایران

خلاصه مقاله:

با توجه به اینکه سازمان دهی اطلاعات آنلاین در قالب دسته بندی مستندات زبان طبیعی در دسته هایی که قبلا بر اساس محتوایشان دسته بندی شده اند یکی از روش های مهم مدیریت اطلاعات محسوب می شوند و با توجه به اقداماتی که در این زمینه برای زبان های دیگر انجام گرفته است، نیاز به طبقه بندی متون فارسی و اهمیت آن به خوبی احساس می شود. روش ها و الگوریتم های متفاوتی برای دسته بندی متن بکار برده شده است، که این روش ها در دقت و محاسبات متفاوت می باشند با توجه به فعالیت های محدودی که در زمینه دسته بندی متون فارسی انجام شده است، در این مقاله به کاوشی بر روی دو روش کلاسه بندی مبتنی بر درخت تصمیم گیری طبقه بندی از نزدیکترین همسایه و مکانیزم عملکرد آنها برای کلاسه بندی اسناد موجود بر روی متون فارسی پرداخته می شود. طبق مطالعات انجام شده روش K تا همسایه نزدیک در عمل کارایی خوبی دارد. مشکل آن اینست که در طی دسته بندی محاسبات زیادی لازم است. الگوریتم های درخت های تصمیم گیری سریع و همچنین مقیاس پذیر هم در متغیرها و هم سایز مجموعه آموزش هستند. یکی از مشکلات درختسهای تصمیم گیری برای متن کاوی اینست که تنها به تعداد کمی از لغت ها وابسته است.

کلمات کلیدی:

کلاس بندی، درخت تصمیم گیری، K نزدیکترین همسایه، پیش پردازش

لینک ثابت مقاله در پایگاه سیویلیکا:

<https://civilica.com/doc/239061>

