

عنوان مقاله:

کاربرد زنجیره مارکوف در ارزیابی و شناسایی شباهت دو فایل متنی

محل انتشار:

همایش ملی کاربرد سیستم های هوشمند (محاسبات نرم) در علوم و صنایع (سال: 1392)

تعداد صفحات اصل مقاله: 7

نویسندگان:

ولی الله سرلک - دانشجوی کارشناسی ارشد مهندسی کامپیوتر، هوش مصنوعی، دانشگاه بین الم

میترا گودرزی - کارشناس مهندسی فناوری اطلاعات، دانشگاه پیام نور الیگودرز

میر مرسل مدنی - عضو هیئت علمی دانشگاه آزاد اسلامی واحد گرگان مرکز کردکوی

سپیده سلطانی فهرج - دانشجوی کارشناسی ارشد مهندسی کامپیوتر، هوش مصنوعی، دانشگاه بین الم

خلاصه مقاله:

در این مقاله پژوهشی هدف مقایسه و ارزیابی و محاسبه درصد شباهت دو فایل از نوع متنی با استفاده از زنجیره های مارکوف بررسی می شود که روش بکار گرفته شده تازه و ایده ای است. متد و راه حل کلی شباهت و ارزیابی دو فایل متنی بدین صورت است که از هر فایل یک بردار شاخص استخراج می گردد، سپس با ادغام دو برداری شاخص، بردار شاخص مرجع یا اصلی ایجاد گردد. با استفاده از بردار شاخص مرجع برای هر فایل یک ماتریس گذار ایجاد می شود. نهایتاً با مقایسه و بررسی دو ماتریس گذار تفاوت هر دو فایل تعیین شده و درصد شباهت آنها محاسبه می گردد. دقت و صحت و درصد شباهت این روش و متد بسیار بالا می باشد و تمام کاراکترهای مورد استفاده در هر فایل (از جمله Space و Enter و Tab) در مقایسه در نظر گرفته می شوند.

کلمات کلیدی:

ادغام فایل، زنجیره مارکوف، مقایسه فایل های متنی، بردار شاخص، ماتریس گذار، Bigram N-gram کاراکتر

لینک ثابت مقاله در پایگاه سیویلیکا:

<https://civilica.com/doc/206214>

