

عنوان مقاله:

روشی جدید در یافتن اسناد یکسان در وب

محل انتشار:

همایش ملی مهندسی رایانه و مدیریت فناوری اطلاعات (سال: 1393)

تعداد صفحات اصل مقاله: 11

نویسندگان:

فریدین ابدالی محمدی - استادیار دانشگاه رازی کرمانشاه

نصرالدین نیازی - دانشجوی کارشناسی ارشد

عبدالحسین فتحی - استادیار دانشگاه رازی کرمانشاه

خلاصه مقاله:

با رشد و گسترش روزافزون شبکه اینترنت و تولید اسناد و فایل های تکراری یا مشابه، سیستم های ذخیره سازی داده ها با مشکلاتی روبرو شده اند. در راستای حل این مشکلات، الگوریتم های زیادی برای تشخیص سندهای تکراری به وجود آمده اند که با حذف سندهای اضافی، تنها یک نسخه ی منحصر به فرد را نگهداری میکنند. با این عمل، مقدار فضای مورد نیاز برای ذخیره سازی کاهش می یابد. همچنین، نقل و انتقال سندهای غیرتکراری در بستر اینترنت به پهنای باند کمتری نیاز خواهد داشت. در این مقاله سوالات جدیدی در مورد روش محاسبه ی مفهومی به نام Containment Document پیش می آید که با پاسخ به آنها و شبیه سازی الگوریتم های پیشنهادی، اثبات خواهیم کرد که میتوان در مدت زمان بسیار کمتری تکراری بودن فایل ها و سندهای موجود در شبکه جهانی وب را نشان داد. به کمک این روش سندهای مشابه در یک خوشه قرار داده میشوند و در مواردی از جمله فیلترکردن نتایج به دست آمده از موتورهای جستجو، به روزرسانی گسترده ی صفحات وب توزیع شده و ... از آن استفاده می شود.

کلمات کلیدی:

شباهت اسناد، شینگل، ابر شینگل، اسناد زیر مجموعه

لینک ثابت مقاله در پایگاه سیویلیکا:

<https://civilica.com/doc/282980>

