

مقایسه ساده الگوریتم های فشرده سازی داده (نسخه PDF)

در این مقاله می خواهیم مروری داشته باشیم بر روش های فشرده سازی و مقایسه آنها در علوم کامپیوتری بحث فشرده سازی به عملی اطلاق می شود که در طی آن داده ها برای کم کردن حجم فشرده می شوند به طور کلی روش های فشرده سازی به دو قسمت تقسیم میشود فشرده سازی بدون اتلاف داده lossless data compression و فشرده سازی با اتلاف داده Lossy compression در فشرده سازی با اتلاف داده بخشی از داده های موجود صرف نظر شده و تلاش می شود تا حد ممکن حجم فایل کاهش یابد که بیشتر از این روش برای صرفه جویی در فضای ذخیره سازی و یا ارسال اطلاعات استفاده می شود و بیشترین مورد استفاده برای فشرده سازی مولتی مدیا تصاویر می باشد .

در مقابل روش فشرده سازی بدون اتلاف داده از متدهایی استفاده می شود تا تمام داده هایی که فشرده سازی شده است بدون هیچگونه آسیب و یا اتلافی دوباره بازیابی شود این روش در موارد استفاده می شود که نیاز است داده های فشرده شده همانند زمانی که فشرده نبوده اند بدون هیچ کم و کاستی بازیابی شوند و هیچ گونه اتلافی صورت نگیرد تمامی برنامه های فشرده سازی بدون اتلاف داده این کار را در دو مرحله کلی انجام میدهند: قدم اول تولید یک مدل آماری برای داده های وارد شده و مرحله دوم استفاده از این مدل برای آدرس دهی داده های وارد شده اساس این نوع فشرده سازی از مدل هافمن کد بهره میبرد ، هافمن کد الگوریتمی است که در این الگوریتم میزان تکرار داده به حداقل رسیده است



یکی از معروفترین فرمت هایی که در فشرده سازی بدون اتلاف داده موجود هست فرمت LZ۷۷ و LZ۷۸ می باشد این فرمتها توسط آقای آبراهام لمپل و جاکوب زیو در سال ۱۹۷۷ ارائه شده اند که امروزه نیز آنها با نام LZ۱ و LZ۲ معروف می باشند و همچنین پایه و بنای بسیاری از فرمت های جدید دیگری نیز می باشند که به وجود آمده اند یکی از دلایل موفقیت این روش کاهش میزان تکرار داده ها بوده است و استفاده از هافمن کد بوده است به طوری که زمانی که شما از این روش برای فشرده سازی استفاده می کنید .

این الگوریتم تمامی داده ها را بررسی و داده های تکراری را شناسایی میکند و صرفاً یک نوع داده موجود را نگه میدارد تا به عنوان ایندکس یا مرجع از آن بهره بگیرد . به طوری که در صورت نیاز به ریکاوری مواردی که در نظر گرفته نشده اند در موقع بازیابی به ایندکس مورد نظر مراجعه و در محل خود که در جدول آدرس دهی مشخص است قرار خواهند گرفت . در مدل پیشرفته این الگوریتم یعنی LZ۷۸ دیگر نیازی به ایندکس دهی نبود در این روش تمامی داده هایی تکراری به دیکشنری موجود ریفرنس داده میشوند و در موقع بازیابی از طریق آن دیکشنری بازیابی می شود.

نویسنده : علی قلعه بان – دانشجوی دکتری فناوری اطلاعات

#بهترین_نرم_افزار_فشرده_سازی_#هافمن_#فشرده_سازی

مطلب اصلی