

بررسی کلی و مقایسه‌ای از مهمترین الگوریتمهای کاوش مجموعه اقلام تکرارشونده بسته

مریم شکفته^{۹۷}؛ امیر مسعود رحمانی^{۹۸}؛ ماشاء الله عباسی^{۹۹}

چکیده

کاوش قوانین وابستگی که یکی از مهمترین روشهای داده‌کاوی است، به کاوش وابستگی‌ها و ارتباطات پنهان بین داده‌ها در پایگاه‌داده‌های بزرگ می‌پردازد و مهمترین عمل در آن یافتن اقلام تکرارشونده است. اقلام تکرارشونده بسته یک روش خلاصه مهم برای اقلام تکرارشونده است که قدرت یکسانی با کاوش کامل مجموعه قلم‌های تکرارشونده دارد، ضمن اینکه قوانین اضافی را کاهش داده و تأثیر کاوش را افزایش می‌دهد. در سالهای اخیر الگوریتمهای زیادی در این زمینه مطرح شده است که از استراتژی‌های مختلفی مانند جستجوی اول عمقی در مقابل جستجوی اول سطحی، قالب عمودی در مقابل قالب افقی، ساختار درختی در مقابل دیگر ساختارها، پیمایش بالا به پایین در مقابل پیمایش پایین به بالا و غیره استفاده می‌کنند. در این مقاله مروری کلی و مقایسه‌ای از مهمترین الگوریتم‌های کاوش اقلام تکرارشونده بسته انجام می‌شود. نتایج آزمایشات نشان داده که هر کدام از این الگوریتم‌ها با توجه به استراتژی به کار برده، در پایگاه‌داده‌های مختلف پراکنده و متراکم دارای یکسری نقاط ضعف و قدرت هستند؛ اما در بین آنها، الگوریتم *DCI-Closed* دارای کارایی بهتری نسبت به بقیه است.

کلمات کلیدی:

کاوش قوانین وابستگی، اقلام تکرارشونده بسته

⁹⁷ دانشجوی کارشناسی ارشد از دانشگاه آزاد، واحد علوم و تحقیقات اهواز- maryam_shekofteh@yahoo.com

⁹⁸ استادیار دانشگاه آزاد، واحد علوم و تحقیقات - Rahmani@Sr.iau.ac.ir

⁹⁹ استادیار دانشگاه آزاد، واحد علوم و تحقیقات اهواز - abbasi_masha@yahoo.com