

Efficiënte methode om gecompriemde databank te wijzigen

Sándor Héman van het CWI ontwikkelde een methode om een omvangrijke database te comprimeren waardoor de gegevens veel sneller van de harde schijf naar de rekenkern van een computer worden overgebracht. Ook ontwikkelde hij efficiënte algoritmes om wijzigingen te kunnen aanbrengen in deze gecompriemde data. Hij verdedigt zijn proefschrift hierover op woensdag 28 oktober aan de Vrije Universiteit. Veel organisaties verzamelen tegenwoordig enorme hoeveelheden digitale informatie, en slaan deze op in een databasemanagementsysteem. Deze database wordt continu gewijzigd en aangevuld, maar ook moet op elk moment geanalyseerd kunnen worden. Dit zijn twee verschillende bewerkingen, die verschillende eisen stellen aan de architectuur van het databasesysteem.

Flessenhals

De analyse van data wordt uitgevoerd door de rekenkern van een computer. De data moeten hiervoor vanaf opslagplaats, meestal de harde schijf van een computer, naar deze rekenkern



worden gebracht. Dit transport van schijf naar rekenkern is veelal traag, wat bij analyse van grote hoeveelheden data een flessenhals kan vormen. “In mijn onderzoek heb ik de negatieve gevolgen van zo’n flessenhals gereduceerd, door de omvang van een databank te verkleinen door middel van data compressie,” zegt Héman. “Door een snelle en transparante compressie te hanteren, kan de rekenkern de data-analyse ongehinderd uitvoeren”

Wijzigingen bijhouden

Als er vervolgens wijzigingen in een gecompriemde databank moeten worden aangebracht

is het omslachtig om de data eerst te moeten decomprimeren, wijzigen, hercomprimeren, en weer opnieuw weg te schrijven. Hiertoe heeft Héman technieken ontwikkeld die wijzigingen differentieel bijhouden, zoals errata bij een boek, maar dusdanig dat deze tijdens het lezen direct en efficiënt worden toegepast.

Hémans onderzoek is relevant voor elk domein waar grote hoeveelheden data geanalyseerd dienen te worden. Dit zijn bijvoorbeeld zoekmachines die het hele web indexeren en doorzoeken, wetenschappelijke toepassingen, zoals analyses van astronomische data of genomen, of commerciële toepassingen waarbij de interesses van grote hoeveelheden klanten in kaart moeten worden gebracht.

Het onderzoek van Héman is ook terug te vinden in het Vectorwise databanksysteem. Vectorwise werd mede door Héman opgericht in 2008 als een spin-off bedrijf van het CWI, en in 2011 verkocht aan Actian Corporation, waar het nog altijd een succesvol product is.



Actueel

Robots leren elkaar nieuwe trucjes

De Europese referentie voor PCB-printing: 2.4-tiende editie

www.eurocircuits.com