

dPersp12 – ugeseddel for uge 40

Internetalgoritmer

Ugens program

Mandag 1/10	10.15-12.00	Forelæsning: Internetalgoritmer - Google's PageRank, MapReduce <i>Gerth Stølting Brodal (PBA Auditoriet)</i>
Tirsdag 2/10	9.15-12.00	Øvelser (Open Learning Center)
	12.15-13.00	Forelæsning: MapReduce eksempler <i>Gerth Stølting Brodal (PBA Auditoriet)</i>
	13.15-16.00	Øvelser (Open Learning Center)
Onsdag 3/10	14.15-16.00	Historisk perspektiv <i>Erik Meineche Schmidt (PBA Auditoriet)</i>

Om ugens forelæsninger

Denne uges forelæsninger omhandler teknikker og algoritmer der i høj grad har påvirket vores brug af internettet de seneste år – med fokus på to eksempler.

Internetsøgemaskiner

Med introduktionen af Google i 1998 blev det muligt at søge effektivt blandt de milliarder af websider der findes på internettet. Nogle af de egenskaber der var fundamentet for Google's succes var hastigheden hvormed forespørgsler udføres, at de fleste sider på internettet var indeholdt i Google's indeks, og ikke mindst at de returnerede svar indeholdt de meste relevante sider om det man søger efter. I forelæsningen beskrives de basale idéer der er grundlaget for Google's virkemåde.

MapReduce

Et af fundamentene i Google's succes er at have mange maskiner (størrelsesordenen en million) forbundet i et stort netværk, og ved at effektivt udnytte den massive parallelisme dette giver. Traditionelle software metoder til programmering af parallelle algoritmer medfører dog i praksis et meget stort overhead til håndtering af tilbagevendende problemer af parallel natur. Google introducerede derfor i 2004 en elegant software omgivelse kaldet MapReduce, der måske ikke kan løse alle problemer effektivt parallelt, men tillader en simpel måde at designe parallelle algoritmer, der automatisk håndterer alle aspekter der vedrører udnyttelsen af parallelisme. Der findes en open source variant af MapReduce, Hadoop, som bl.a. bruges intenstivt af Yahoo!