

Applicazioni tecnologiche della teoria dei linguaggi: la parsificazione in parallelo nei browsers e servers

Stefano Crespi Reghizzi

professore emerito del Politecnico di Milano

mercoledì 19 marzo 2014, h. 15:00-16:00

AULA SEMINARI cubo 42C - V piano – DIMES

ABSTRACT

Come sfruttare le capacità di calcolo parallelo presenti in tutti i dispositivi per accelerare l'analisi di testi di grandi dimensioni? Parallelizzando l'elaborazione, si riduce inoltre il consumo di energia elettrica e il calore dissipato, e si prolunga l'autonomia dei dispositivi alimentati a batteria. Purtroppo, l'analisi sintattica o parsing classica è poco parallelizzabile perché l'algoritmo, per prendere una decisione in un certo punto del testo, deve conoscere l'analisi del pezzo precedente; così si impedisce l'analisi in parallelo di più segmenti del testo sorgente. Sotto la spinta delle esigenze dei browser per HTML, da un lato e dei server per le ricerche in rete, dall'altro, sono stati fatti passi avanti per parallelizzare l'analisi sintattica. Due esempi sono considerati. Mozilla Corporation (e altri) ha investito nello sviluppo di parsificatori paralleli per HTML, superando con un ventaglio di tecniche i principali ostacoli che tale linguaggio presenta. Il nostro progetto, finanziato in parte da Google, sfrutta le proprietà di parsabilità locale delle grammatiche a operatori con precedenza (utilizzate nei compilatori per FORTRAN degli anni 1960!), che a tutt'oggi sono la classe più ampia di linguaggi che offrono la possibilità di sfruttare il data-parallelism, grazie alla loro peculiarità di essere "visibilmente a pila" (visibly push-down / input-driven). Lo strumento PAPAGENO, scaricabile dalla rete, genera dei parsificatori paralleli, scritti in C, molto veloci, e scalabili al crescere del numero dei processori fisici; e, per guadagnare ulteriore velocità, parallelizza anche l'analisi lessicale. Esso dà ottime prestazioni per il linguaggio JSON, spesso usato dal lato server per rappresentare dati semistrutturati. Si conclude con l'indicazione degli sviluppi possibili.