

## Keywords

20 - Public service

13 - Research

02 - Algorithms, 06 - Databases

03 - Software (data analysis)

JORDAN Vincent

ST50&AHPM internship report - P2010

## Abstract

*XML* is a machine-readable language designed to be both simple and extendable. It is widely used in computer world for data storage and communication therefore the need of querying XML data is high. The purpose of the research carried out during my stay at Kitagawa Data Engineering laboratory is to evaluate the possibility of using *General Purpose Graphic Processing Unit* (GPGPU) to handle this task.

GPGPU is a recent creation which gather research attention because of its large diffusion and cheap price. GPGPU is the result of the evolution of GPU coprocessors toward more flexibly and programable features in visual rendering. The latest features of these chips enable them to be used as stream processors.

*Stream processing* is a highly data parallelized task in which a series of operation is applied to each element of a data set (a stream).

Research on parallel XML query processing has already be done in the same laboratory by IMAM MACHDI. My main work is to create a GPU algorithm based on his PhD thesis results. This task is challenging since current XML processing algorithms do not fit in the stream processing paradigm.

Two solutions are possible to solve this issue:

- overcome GPU limitations in order to do more than stream processing
- create a new algorithm which is stream processing compliant

*During this internship, the first solution has been studied.*

## Kitagawa Data Engineering laboratory

University of Tsukuba

Tennoudai 1-1-1, Tsukuba, Ibaraki, Japan 305-8573



*University of Tsukuba*