ABSTRACT OF THESIS

Currently, the demand for searching and document sharing to serve for research and teaching at university is enormous. Besides, cheating situations like copying the thesis from previous semesters or even from other universities are happening sophisticatedly and popularly. This negatively affected to the quality of the thesis particularly and the quality of education in general. Therefore, the need for a document sharing system which provide the ability for comparison in anti-fraud studying is very urgent.

However, managing distributed documents between universities is very challenge. They are dynamic resources which are belonging to different organizations and each organization has its own private policy in resource and access management. On the other hand, a distributed document managing system requires flexibility (easily add and remove storage and computing resource), security, and collaboration between research groups. To handle this requirement on a set of dynamic resources is complicated.

Develop from the demand; we propose and deploy a model system of electronic document management based on data grid technology **GOODAS - Grid Oriented Online Document Analysing System**. The system will connect computing resources and storage of dispersed universities for creating a grid inter-university data. On that basis, the system supports managing users, as well as searching document within the field or on the entire system. Especially, the system also provides the ability to match the document content, based on LDA algorithm - an approach to the semantics of analyzing the document content. For each input document, the system will rely on existing data warehouse to analyze the match for their content, and conclude how much the document is copied from other documents. Besides, we also conducted experiments grouped according to the document fields. Results showed that the system has enhanced the capacity to meet as well as the accuracy of the search and matching documents.

My job in the group is building data grid component which manages thesis according to research groups. This component must ensure these propersites of the system: **security**, **reliability**, **availability** and **high scalability**. The system (GOODAS project) is currently deployed for testing at High Performance Computing Center, Hanoi University of Technology.

Keywords *- Grid computing, matching documents, distributed search, grid security*