Carlos André Reis Pinheiro

"Warehousing"

The number of access and the information that flow in the web increaseeveryday. The avaiable and possible extract information grow tax makes theInternet the biggest library in the whole world. This library have othersfeatures, like be present anywhere, besides of the your dynamic content. Nevertheless, its necessary becames the content of the Internet persistent.To order deploy the Internet development into a digital library, otherstools and applications must be created. The existing web tools do not offer a good solution to manage and maintain those information included in Internet. Besides that, the information insertted in the Internet are not formated and grouped to order improve the users´s visualization. In this way, was elaborated the Web Warehousing system concept, that have as a main goal, support activities of contents analysis of the information extracted through the Internet. This kind of system must allows the user extract, archive, query and select segments of the Internet. Those segements of the Internet can be mantained in a Web Warehousing environment to futures operations, like a complex queries, text mining and text summarizing. The goal of this work is to proceeding a analysis of the development methods to build a Web Warehousing. In this way, will be presented yours general concepts, describing the general architecture, the methodology to coupling web informations, the algebra e data model to internet, and the visualization operators to manipulate informations from Web Warehousing. The case study clarify how this methodology can provide proficient in the management the extration and manipulation of data from web, to order support adacemicals researchs, became the internet in a digital library, even to a limit scope directed to a some application. Based on approachs about data warehouse and extraction and manipulation of web informations, which were presented in this work as a Web warehousing environment, the goal is to apply this point of view in a case study focused to Federal University Fluminense, more precisely to the Pos Graduate Program in Computer Science - PGCC/UFF.