

**CHENNAI – PONDICHERRY**

**AUTOMATIC CONSTRUCTION OF VERTICAL SEARCH TOOLS**

**FOR THE DEEP WEB**

**ABSTRACT**

With the constant increase in the volume of information available on the Web, it is more dificult to find the specific information related to a given domain. Users are facing the problem of information overload, in which a query about a specialized subject (local information, e-commerce: hotels, airlines, car rental; science: biology, mathematics, medicine, etc.) on a web search engine, it returns a lot of web pages or results that in most of the cases are outside the domain of interest. This is one reason why the vertical search tools have become a necessity for users that seek specific-domain information from diferent databases available in the Web through input sources called Web Query Interfaces (ICWs). This paper describes an approach for automatic integration of ICWs, a crucial task to construct vertical search tools. The proposed methodology is validated by realizing a vertical search prototype called VSearch that allows users to transparently query multiple web databases in a specific-domain through a unified ICW. The proposed approach for automatic ICWs integration is based on: i) a hierarchical model called AEV for modeling the visual content of ICW; ii) semantic clustering for the identification of relationships between fields in ICWs; and iii) a field homogenization and unification process of AEV schemes for the construction of a unified ICW. The VSearch prototype was implemented and evaluated using a study case. The experimental results demonstrate the high precision in the integration phase and an efective methodology to create a functional vertical search tool for a given domain.