


Chapter 21

The Proposed Framework of View-Dependent Data Integration Architecture


Pradeep Kumar

 <https://orcid.org/0000-0002-0206-9808>
Teerthanker Mahaveer University, India

Madhurendra Kumar

Ministry of Electronics and Information Technology, CDAC, India

Rajeev Kumar

 <https://orcid.org/0000-0002-4141-1282>
Moradabad Institute of Technology, India

ABSTRACT

In this chapter the authors have proposed framework that have we are using various techniques to overcome the above mentioned challenges and proving the exact identification and extraction of WQI by classifying them according to their domain. The framework has been represented in system level design as a high level view of the view dependent data integration system and the architectural framework of the system design. Being a multi-database-oriented system, it is scalable to structure as well as unstructured data source. The wrapper and mediator module of the system is used to map the web-query-interface to the global schema of the integrated web query interfaces. In this the authors have implemented high level view of the system modeling from the end users' point of view; and the operational framework design also been represented.

INTRODUCTION

Deep web provides huge amount of domain-specific content like medical data, houses on sale data, e-commerce and science data etc. All these data stored on deep web can be accessed with the help of various HTML forms named Web Query Interface. This information stored in deep web can be fetched

DOI: 10.4018/979-8-3693-2964-1.ch021

from multiple database servers but one at a time that makes the information accessing process inefficient. To overcome this, a different approach is used by integrating Web Query Interface (IWQI) that is able to perform the query operation on multiple database servers at a time and acts as an independent entry point. In the proposed architecture we have given an innovative method for integrating web forms along with IWQI based on VDIS (View-based Data Integration System) and linked data. We have proposed a novel and alternative solution for combining various WQI into single IWQI for a specific domain for the advanced registrations based systems for enhanced security and reliability. Our proposed method starts with an integrated system by wrapping of various domain-wise WQI till the development of single integrated WQI.

The deep web, also called invisible web, hidden web is a part of World Wide Web. Its content can't directly index by common search engines like Yahoo, Google or Bing. Deep web acts like an umbrella for various parts of internet that are not fully accessible through the mentioned search engines. In the working process of crawling a web, every search engine uses search robots to add new contents to find any appropriate search engine index. The range of deep web is unknown to everyone, but according to the estimates by experts that 1% of all content over the web is indexed and crawled by search engines. The content that is available for access through search engines is known as surface web. As the content stored in it, deep web includes private data over various social media sites, emails, chat messages, bank statements etc. Search Robots are capable to search the paid sites like news articles or educational sites those requires some kind of subscription and for the open access sites like YouTube, Netflix. The terms deep web and dark web are uses similarly sometimes, still they are having some differences. Dark web is a part of larger web and the content stored in it is accessible via any search engine like Google.

Dark web is less reputed as it involves many illegal activities like malware attacks, cyber attacks, black market of stolen credit card, etc. In contrast of this, deep web contains legal content like academic journals and research databases. Although deep web can't be indexed by standard search engines, still, to access deep web is safe and easily accessible by many users. Gmail, LinkedIn sign in are some of the examples of deep web sites. The data on deep web contains personal information, that's why they are having restrictive access. The value of deep web content becomes enormous if the most iconic commodity of current information age is actually information. By having these facts, Bright Planet shows various facts and findings in a study. The study is actually based on collected data between 13 and 20 March 2000. The findings include-

- Deep web is having a huge public information dataset. There available information is 400-500 times larger than the data in commonly defined in www.
- Deep web is having 7500 Tb of information in its database. On contrary surface web is having only 19 Tb of information.
- 550 billion individual documents are stored in deep web that are far more than surface web.
- In current time near about 200,000 deep websites exists all over the web.
- On an average, data traffic received by the deep websites is 50% more than the data traffic received by the surface web.
- Content available on deep web is highly connected and relevant to any information related to market, need or domain.
- Deep web provides quality content 1000- 2000 times greater than surface web.

17 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/the-proposed-framework-of-view-dependent-data-integration-architecture/341203

Related Content

The Potential for ICT Tools to Promote Public Participation in Fighting Corruption

Arjun Neupane, Jeffrey Soar, Kishor Vaidya and Sunil Aryal (2015). *Business Law and Ethics: Concepts, Methodologies, Tools, and Applications* (pp. 338-354).

www.irma-international.org/chapter/the-potential-for-ict-tools-to-promote-public-participation-in-fighting-corruption/125740

Guest Socioeconomic Status and Hotel Green Technology: Manager Entrepreneurial Advantage

Faranak Memarzadehand Sulekha Anand (2020). *International Journal of Sustainable Entrepreneurship and Corporate Social Responsibility* (pp. 55-65).

www.irma-international.org/article/guest-socioeconomic-status-and-hotel-green-technology/259408

Scaling Sustainability Value in Sustainability Purpose Organizations: A Non-Profit and Business Comparison

Alex Lyakhov, Travis Gliedt and Nathan Jackson (2016). *International Journal of Sustainable Entrepreneurship and Corporate Social Responsibility* (pp. 17-31).

www.irma-international.org/article/scaling-sustainability-value-in-sustainability-purpose-organizations/172165

Guest Socioeconomic Status and Hotel Green Technology: Manager Entrepreneurial Advantage

Faranak Memarzadehand Sulekha Anand (2020). *International Journal of Sustainable Entrepreneurship and Corporate Social Responsibility* (pp. 55-65).

www.irma-international.org/article/guest-socioeconomic-status-and-hotel-green-technology/259408

Factors Affecting Listening

Vehbi Turel (2014). *Human Rights and the Impact of ICT in the Public Sphere: Participation, Democracy, and Political Autonomy* (pp. 310-331).

www.irma-international.org/chapter/factors-affecting-listening/112182