Application of VIVO Semantic Web Technology in Cadre Information Management

Authors: Ren Wei, Huang Jinxia

Date: 2016-06-12T00:00:00+00:00

Abstract

Abstract: The cadre information management system suffers from issues of low data management efficiency and simplistic search results. Based on the implementation mechanisms of VIVO semantic web technology in linked data, ontology knowledge bases, and semantic search, this paper designs an intelligent analysis system architecture for cadre information, constructs a cadre information ontology knowledge base defined by the RDF protocol, establishes relationship mappings between different types of data, realizes data association and implicit data discovery, and enhances the semantic search efficiency of the cadre information management system. The structured and associative ontology framework also reduces information management costs.

Full Text

Application of VIVO Semantic Web Technology in Cadre Information Management

Ren Wei[1], Huang Jin-xia[2]
(1. College of Engineering and Computer Science, Australian National University, Canberra, Australia;
2. National Science Library, Chinese Academy of Sciences, Beijing 100190, China)

Abstract

Cadre information management systems suffer from inefficient data management and simplistic search capabilities. Drawing upon VIVO semantic web technologies for linked data, ontology-based knowledge bases, and semantic search implementation mechanisms, this paper designs an intelligent analysis system architecture for cadre information. The system constructs an ontology knowledge base for cadre information using RDF protocols, establishes relationship mappings between different data types, achieves data linking and hidden data discovery, and improves semantic search efficiency in cadre information management. The structured and interlinked ontology framework also reduces information management costs.

Keywords: VIVO; Semantic Web Technology; Ontology-based Knowledge Base; Cadre Information Management System
Chinese Library Classification: TP31
Document Code: A

Cadre management information systems occupy a central position in organizational and personnel management. With the scientific development of cadre management, the informatization of cadre materials—including digital storage, information management, and decision support for cadre selection—has become a critical challenge for organizational departments [1-2]. Currently, most cadre information management systems only provide basic display and simple retrieval functions, failing to meet the requirements for scientific cadre management [3]. There is an urgent need to employ advanced information technologies to establish secure, efficient, accurate, and convenient cadre information management processes. Such systems would not only make cadre information management more scientific and convenient but also provide valuable support for cadre selection, appointment, and training.

Search engines worldwide continue to face challenges with insufficient recall and precision. Although Google began researching semantic technologies early on, it still struggles to meet the specific semantic retrieval needs of specialized users [4]. Semantic search technology can improve search effectiveness, yet current research faces two major obstacles: the lack of effective methods for parsing user semantics and the absence of a mature semantic search model [5]. Due to variations in data volume and application contexts, no "universal" solution exists. Grounded in situational theory and based on ontological methodology, using ontologies to express abstract situational factors as concrete variables that information retrieval systems can process may evolve into an ontology-based semantic information search model [6].

The U.S. VIVO project employs Semantic Web technologies such as RDF, OWL, and linked data to construct ontology-based semantic tools [7]. These have been successfully applied in diverse contexts, including the VIVO research network across over 20 U.S. universities [8], the USDA' s information resource discovery system [9], and the Subject Knowledge Environment (SKE) for more than 100 institutes of the Chinese Academy of Sciences [10]. These applications support structured and linked processing of information to enable semantic search across multiple sources. This paper addresses challenges in the scientific management and utilization of cadre information by leveraging VIVO semantic web technologies and search models to construct an intelligent analysis platform that improves the efficiency of cadre information management and utilization.

2.1 VIVO Overview

Semantic search integrates Semantic Web technologies into search engines to improve effectiveness. It can be categorized into enhanced semantic search based on traditional search methods and knowledge-based semantic search grounded in ontology reasoning [5]. The latter, which uses ontology-constructed knowledge bases to achieve knowledge discovery through reasoning, has been successfully demonstrated by the VIVO project.

VIVO (VIVO Is a Virtual Organization) was initially launched by Cornell University Library in 2004 to support life sciences research in the College of Agriculture and Life Sciences. In 2007, VIVO was re-engineered using RDF, OWL, Jena, and SPARQL technologies to enable semantic retrieval and association discovery for faculty, researchers, and disciplines across all Cornell colleges. Building on VIVO' s success at Cornell and the need for research collaboration, seven U.S. universities jointly secured $12.2 million in funding from the National Center for Research Resources (NCRR) of the National Institutes of Health (NIH) in September 2009. This funding supported the development of an open semantic application designed to facilitate networked collaboration among researchers. VIVO' s software and tools are open source, enabling other developers with similar needs to rapidly and effectively build applications that realize VIVO' s vision of creating national and international scientific networks [11].

2.2 VIVO Core Technologies

VIVO employs Semantic Web technologies including linked data and ontology technologies, which are widely used and will not be elaborated here. Notably, VIVO utilizes these technologies in a straightforward manner to structure and link multi-source, multi-type information, thereby supporting semantic and associative retrieval for users while enabling efficient underlying data management for system administrators. The core of VIVO is its ontology model and knowledge base framework, for which it developed the Vitro ontology editor. Vitro enables VIVO users to rapidly create ontologies and knowledge bases, implement retrieval functions, and support import, export, and SPARQL queries of RDF data. Additionally, Vitro provides services for flexibly linking display content with the underlying ontology knowledge base, supporting rapid user platform development [12].

2.2.1 RDF Framework and Linked Data Technology

Resource Description Framework (RDF) is a language for expressing information about resources on the World Wide Web. Its fundamental concept is that all information is considered a "resource" with "properties" that have "values," with resources described through "statements." RDF uses triple statements <subject, predicate, object> to describe Web resources [13]. Linked data is a practical Semantic Web expression that employs the RDF data model and uses URIs to name data entities for publishing and deploying class and instance data. This enables different sources to reveal and obtain data via HTTP while emphasizing data interlinking and contextual information understandable by both humans and machines. VIVO leverages linked data technology to discover associations among data distributed across member institutions' VIVO systems, thereby creating scientific networks.

2.2.2 Ontology and Ontology Knowledge Base

Ontology is an explicit and formal specification of a shared conceptualization [14]. Shared concepts include conceptual frameworks for modeling domain knowledge, content-related protocols for interaction between interoperating agents, and common agreements for representing domain-specific theories. A knowledge base establishes well-structured knowledge models that conveniently describe and organize related knowledge. While knowledge models can represent knowledge as text, textual knowledge is not conducive to processing operations such as reasoning and classification. Ontology-based knowledge models can support data classification and implicit knowledge reasoning [15]. The VIVO ontology is the core component of VIVO development. Its design strategy involves building a generic core ontology and local ontologies with master-slave relationships. Ontologies in different namespaces support queries of all or local ontology instances. Based on the ontology model, external information is ingested into VIVO and matched with ontology classes and properties to achieve structured processing, ultimately forming instances that enter the core or local ontology knowledge base. The instance generation process is complex; VIVO provides specialized data ingestion tools, as described in literature [16].

3 Design of the VIVO-Based Cadre Information Intelligent Analysis Platform

The cadre information intelligent analysis platform develops cadre information management functions based on semantic search concepts. Through semantic search, computers can accurately comprehend the true meaning of user queries and conduct searches accordingly, precisely returning the most relevant results to achieve effective information management. This function is implemented through an RDF protocol-based ontology knowledge base for cadre information management and search, providing preliminary analytical capabilities such as tenure experience analysis to meet specialized departmental needs.

Using VIVO's Vitro software as the platform foundation and adopting the VIVO core ontology to build the knowledge base, the platform defines and describes terms to reduce ambiguity. To improve retrieval efficiency, new inference rules and search strategies are formulated for efficient information resource reasoning. Additionally, since ontology organizational structures differ significantly from traditional data in existing cadre information management systems, data conversion between the intelligent platform and traditional platforms—specifically, effectively transforming traditional data into formats required by the ontology knowledge base—represents a major challenge in platform construction.

3.1 System Architecture

The cadre information intelligent analysis platform provides semantic search of cadre information, growth experience analysis, relationship networks for searched individuals, and knowledge base management, data management, and account management for personnel administrators. The system architecture is shown in Figure 1 [Figure 1: see original paper]. The modular design comprises two main components: the user application function module and the development and maintenance function module. The user module enables semantic searches by cadre name, position, and other criteria; visualization of growth trajectories analyzed from "specific experiences" ; and discovery of relationship networks inferred through rules. The development and maintenance module supports platform administrators in managing information, including creating and modifying the cadre ontology knowledge base, importing and exporting data, and implementing access permission management for different user levels (e.g., leaders, internal personnel, general visitors).

3.2 Key Technologies

Platform construction involves three primary challenges: cadre information ontology construction, cadre information ontology knowledge base construction, and semantic search model construction.

3.2.1 Cadre Information Ontology Construction The cadre information ontology reuses the VIVO core ontology, which defines 13 top-level classes covering people, organizations, academic activities, publications, space, time, and other domains, with diverse properties for each class. For example, the secondary class "Management Personnel" under "People" includes properties such as name, educational experience, work experience, published papers, published works, participated projects, and received honors [17].

3.2.2 Cadre Information Ontology Knowledge Base Construction The cadre information ontology knowledge base is constructed using VIVO' s applications and related tools. Data sources include existing traditional databases and self-input from platform administrators or users. The former requires batch data ingestion by system administrators, while the latter requires users to describe information according to ontology instances.

Data ingestion into the ontology knowledge base requires first categorizing data from traditional databases, then converting relational data to XML, and subsequently XML to RDF, ultimately forming the ontology knowledge base [18-19].

Step 1: Convert original cadre information data into XML format. The program structure is as follows:

Step 2: Convert XML into RDF format to establish a complete knowledge base. The program structure is as follows:

3.2.3 Semantic Search Model Construction When users perform searches, the system processes keyword-based queries to provide traditional search results of instance status. Simultaneously, using association rules and inference rules defined in the ontology (i.e., the inference engine), the system infers and discovers additional associated instances from the ontology knowledge base based on retrieved instances, thereby discovering implicit information. The user query requests and retrieval processing flow are illustrated in Figure 2 [Figure 2: see original paper].

3.3 Implementation Details

3.3.1 Database Configuration The system must export basic data and data dictionaries from traditional or storage databases, making basic database system construction essential. The cadre information intelligent analysis platform requires complex data conversion, and database configuration is necessary after framework establishment before actual system implementation.

1. Create Map to obtain data dictionary

```
public class StandardExcelToMap {
    public static Map<String, Map<String,String>> readExcel(File file) throws IOException {
```

2. Convert original database data to XML

```
public class DataProcessiongSolution1 {
```

3.3.2 Knowledge Base Module Implementation Knowledge base implementation forms the system foundation. The process involves converting XML to RDF, creating corresponding ontologies based on database structure, and finally building the knowledge base according to the created ontologies.

1. Import required classes and related reasoners, and name them in the VIVO space

```
public class KnowledgeBase_new {
    private static String nameSpace = "http://www.semanticweb.org/whj/ontologies/2013/4/gove
```

3. Place prepared instances into Map according to data dictionary

```
private static void initialize() {
```

4. Establish cadre resume knowledge base according to ontology

```
public void addResumeKnowledge(File file) {
```

5. Create new cadre resumes based on instances in ontology

```
private void recurCreate(Element el, Individual parentIndividual,long startTime, int level)
```

6. Match cadre resumes with instances in ontology

```
public Individual entityResolution(OntClass elClass, String entityName) {// Parse correspond
```

7. Use Jena reasoner to reason ontology

```
Public Map<String, String> queryIndividualByClass(OntClass elClass) {// Search by class name
```

8. Store organization instances in knowledge base

```
public void addSchoolKnowledge(OntModel model1) { // Create knowledge base in OntModel, add
```

3.3.3 System User Management Based on different levels of cadre information management and application, the system provides separate management for ordinary users and administrators. Ordinary users log in to the main interface and can: perform semantic searches (by keyword or name with constraints such as ethnicity, political affiliation, gender, position level, and birth date, with further filtering by highest education level in results); conduct relationship analysis (displaying marital relationships, direct relatives, collateral relatives, and close affines); and analyze growth experiences, birthplaces, cadre evaluations, and team compositions. Administrators, after logging into the main interface, can: maintain the knowledge base (query existing ontologies, manage class hierarchies and groupings, control object and data property hierarchies, and edit/display SPARQL statements); manage data (add/delete RDF data, download extraction tools, and export RDF data); and manage users (add, delete, or query users).

4 Implementation of Core Functions

4.1 Semantic Search Function

The search module serves as the core frontend component, supporting primary user needs. Users search for personnel through the search bar, and the displays both conventional and semantic search results. Beyond basic user information in conventional results, the semantic search content vividly presents deep-level information such as interpersonal relationship networks, tenure experiences, and family member identities through dynamic tables and graphics.

4.2 Relationship Analysis Function

Relationship analysis enhances understanding of individual cadres and represents a key focus of cadre information management. Platform users can view specific interpersonal relationships of relevant personnel, enabling more direct and effective management of position changes.

4.3 Growth Experience Analysis Function

Cadre growth experience analysis is another critical platform function. It displays searched individuals' growth trajectories through intuitive charts, including: implementation of growth experience search functions; composition of growth experience analysis; and tabulation of personnel experiences to achieve final visualization.

4.4 Knowledge Base Maintenance Function

This function facilitates knowledge base maintenance by administrators and developers, including: ontology editing, import, and export; control of ontology class hierarchies, object properties, and data properties; data management functions for basic CRUD operations; and user management functions for adding, deleting, and querying users, as well as enabling users to manage their own accounts.

Conclusion

Semantic search services are positioned in knowledge-intensive domains, providing systematic, comprehensive, and specialized answers to professional questions rather than the general services required by traditional search engine users. The cadre information management platform constructed in this paper is an intelligent information analysis platform built upon ontology knowledge base concepts, VIVO semantic web technologies, and SPARQL query language. It offers practical solutions for meeting the informatization and scientification requirements of current cadre management, distinguishing it from existing cadre information management systems that typically focus on architecture without addressing data organization or emphasize information management without considering user utilization. Leveraging VIVO' s technologies and tools, the platform offers high maintainability and operability, facilitating development, maintenance, and management while reducing economic investment in cadre information management and application. Current VIVO tools provide significant value in linked data; enhancing semantic enrichment tools would further promote intelligence in information management and application [20].

References

[1] Wang Ying-long, Pan Yu-chun. Design and Implementation of Cadre Management Information System Based on ASP.NET MVC [D]. College of Computer and Information Engineering, Jiangxi Agricultural University, 2014.
[2] Liu Wei. Preliminary Exploration of the Design of Comprehensive Evaluation Information Management System for University Department-Level Reserve Cadres [J]. Manager, 2015(6):
[3] Du Qing-tao. Design and Implementation of Cadre Management System [D]. College of Computer Science and Technology, Jilin University, 2015.
[4] Amerland D. Google Semantic Search: Search Engine Optimization (SEO) Techniques That Get Your Company More Traffic, Increase Brand Impact, and Amplify Your Online Presence [M]. Que Publishing, 2013.
[5] Wen Kun-mei, Lu Zheng-ding, Sun Xiao-lin, et al. Survey on Semantic

Search [J]. Computer Science, 2008,35(5):1-4.
[6] Xu Jing, Sun Tan, Huang Fei-yan. Progress in Foreign Ontology Application Research in Recent Two Years [J]. Library Development, 2008(8):84-90.
[7] Katy Borner, et al. VIVO: a semantic approach to scholarly networking and discovery [M]. San Rafael, Calif. Morgan & Claypool, 2012.
[8] VIVO: Research & Expertise Across Cornell [EB/OL]. (2015-11-30) http://vivo.cornell.edu.
[9] VIVO: USDA Science & Collaboration [EB/OL]. (2015-11-30) http://vivo.usda.gov.
[10] Song Wen, Huang Jin-xia, Liu Yi, Tang Yi-jie. SKE Key Technologies and Services for Knowledge Discovery [J]. New Technology of Library and Information Service, 2012,7/8:13-18.
[11] Krafft D B, Cappadon N A, Devare B M, et al. VIVO: Enabling National Networking of Scientists [EB/OL]. (2015-11-30) https://www.researchgate.net/publication/228564370_VIVO_
_of_Scientists.
[12] Huang Jin-xia, Song Wen, Liu Yi, et al. Construction and Application of Subject Knowledge Environment of Chinese Academy of Sciences [J]. Library and Information Service (Supplement 2), 2010:337-341,352.
[13] Tong Qiang, Zhang Fu, Cheng Jing-wei, et al. Research on RDF Storage in Relational Database [J]. Journal of Northeastern University (Natural Science), 2015,36(3): 346-349.
[14] Gruber T. Toward principles for the design of ontologies used for knowledge sharing [J]. International Journal of Human and Computer Studies, 43(5/6):907-928.
[15] Yuan Lei, Zhang Hao, Chen Jing, et al. Research on Knowledge Base Construction Pattern Based on Ontology Knowledge Model [J]. Computer Engineering and Applications, 2006(30):65-68,104.
[16] Huang Jin-xia, Jing Li. Data Ingestion Tools for VIVO Ontology [J]. New Technology of Library and Information Service, 2011,22(1):16-20.
[17] Mitchell S, Chen S S, Ahmed M, et al. The VIVO Ontology: Enabling Networking of Scientists [EB/OL]. (2015-22-30) https://www.researchgate.net/publication/228934646_The_VIVC
orking_of_Scientists.
[18] Li Wen. Research and Application Analysis of Web Information Extraction Technology [J]. Electronic Technology & Software Engineering, 2015(3):15.
[19] He Shao-peng, Li Jian-hui, Shen Zhi-hong, et al. Survey on Large-Scale RDF Data Storage Technology [J]. Network New Media Technology, 2013,2(1):8-16.
[20] Huang Jin-xia. Case Study on Semantic Web Applications Supporting Scientific Research and Academic Discovery [J]. Library and Information Service, 2011,55(6):126-129.

Author Biographies

Ren Wei (1991-), male, from Beijing, master' s student at the College of Engineering and Computer Science, Australian National University. Research interests: data analysis and semantic search.

Huang Jin-xia (1972-), female, from Tianchang, Anhui, Ph.D., master' s supervisor, associate research librarian. Research interests: domain ontology construction, data utilization of open scientific resources.

Note: Figure translations are in progress. See original paper for figures.