

Managing Social Security Data in the Web 2.0 Era

Li Luo¹, Hongyan Yang², Xuhui Li³

¹College of Literature, Law & Economics, Wuhan University of Science and Technology, Wuhan, China; ²School of Political Science & Public Administration, Wuhan University, Wuhan, China; ³State Key Lab of Software Engineering, Wuhan University, Wuhan, China.

Email: moli0913@163.com, yhyhyang@163.com, lixuhui@whu.edu.cn

Received May 23rd, 2012; revised June 23rd, 2012; accepted July 23rd, 2012

ABSTRACT

Social security data management is an important topic both in application of information management and in social security management. In the Web 2.0 era, more and more human information and healthcare information is released to the Internet through various approaches. This abundance makes managing social security data go beyond managing conventional social security database records. How to organize the conventional records together with the related information gathered from the Web is an interesting problem to solve to provide more convenient and powerful social security information service. In this paper, we introduce our initial work on building a Web-oriented social security information system named i-SSIS. I-SSIS is a database system which adopts a new object-role data model named INM model and deploys INM database system as its core. With the assistance of auxiliary tools to carry out social security information extraction, analyzing and query, i-SSIS can properly provide social security-related information gathered from the Web. We introduce the basic ideas of designing i-SSIS and describe the architecture and major components of the system.

Keywords: Human Resource Management; Social Security; Data Management; Information System

1. Introduction

Managing social security data has a long history in data management, especially in e-Government. In the early days of data management [1], US government has designed a well-defined database to manage social security data and dwelled on the major issues of building such a system. Although social security data management involved many practical problems at that time, it was still a typical application of database system or, more precisely, management information system where a well-designed database played a central role and the key technical problems were to organize the data and to design the queries for common applications.

When the Web 2.0 era comes, lots of things in information management are changed. There is an abundance of data in all fields, and the approaches of providing, sharing and utilizing the data has varied a lot. In the field of social security data management, the technical key point is migrating from organizing data and queries to gathering and utilizing data. That is, although the central component here is still the social security records, a lot of information in the Web related to people and other social security issues is available and can be utilized together with the social security records. More knowledge can be discovered in the combination of them, and be further

used for governments and organizations in making decision. For example, as an important portion of social security data, health care information was usually managed in the same way as other kinds of data. However, in practical applications, it is fairly common for people to know the associations among the specific group of people, the diseases, the therapies and the expenses. The information here involves not only the pure social security records, but also the data about people's career background, the clinical information, and so on. The extra information was hard to get in the pre-Internet era, however, it is not difficult to gather and analyze in the Web 2.0 era since lots of people are sharing their personal information in blogs and various kinds of medical information can be explored and gathered from the Web, e.g., Wiki. Managing social security data now seems going beyond managing social security records. It is becoming a task of integrating and utilizing various kinds of data involved in human and social security and providing various information services for the people who concerns issues related to social security. In this task the Web plays an important role.

To accomplish the task mentioned above, data management approaches and tools need to be improved and enhanced. However, up to now, there is still a lack of study considering the challenges brought about by the

new task, and seldom researchers in the fields of data management and social security propose a feasible and practical way of establishing a scheme to provide the social security information service which embodies the features of the Web 2.0 era.

In this paper, we propose a framework of a new social security information system for the Web 2.0 era. In this system named i-SSIS standing for Internet-oriented Social Security Information System, data management of social security is no longer based on a conventional database. We deploy a new database system that can manage the human resource information and other information related to social security. Based on this novel system, many new value-added information services for social security can be developed, and among them the social security search service is what we are devoting to build.

The rest of the paper is organized as follows. In Section 2 the related works about social security data management and utilization is briefly discussed. Section 3 illustrates a scenario indicating how the information sources in the Web affect the social security. In Section 4, the architecture of i-SSIS is illustrated and the major components are introduced. In Section 5, the ongoing work on searching service of social security information is introduced. Section 6 concludes the paper.

2. Related Works

Research on social security data management and utilization can trace back to the late of 1970s [1,2] and keep progress with the development of the studies in both social security and information technology. Nowadays, government departments and companies involving the affairs of social security all have their databases and information systems of social security. Various kinds of studies in the information technology fields have been worked on social security data.

Some studies on social security data mining try to analyze the social security data to find patterns of social security related affairs, such as debt [3,4] and health care [5]. Some studies concentrate on protecting social security data [6] since they are often confidential and should be accessed only for specific use.

Although social data are usually stored and managed by conventional database system, there is a new tendency of developing information management for social security data to cater for current technology of data utilization such as data mining. For example, studies are working on developing data warehouse for healthcare data for efficient mining [7].

Another relevant field with rapid development is human resource management [8] where people are building more comprehensive and efficient information system to collect and manage various kinds of human resource information.

Inspired by these trends in social security data management and utilization, we propose a new information system to combine the human resource information and conventional security data and thus can be used for various kinds of people, e.g., managers, researchers and officials, to access and utilize the data with specified authorizations and privileges. This system named i-SSIS is oriented to Web 2.0 resources from which we can gather the information of people directly and link the relevant data to form a rich information network. By integrating social security data with the human resource information gather from the Web, various kinds of data analysis can be carried out to provide rich social security information services.

3. A Scenario

In conventional social security data management, the database stores the basic information of each person, such as the career and education background, healthcare records, salary and pension records, etc. These records can provide important information for administrators or researchers to make some basic statistics and decisions. However, if we want to know more about the relationships between certain information involved in social security, conventional records are often not enough.

For example, a researcher in social security field wants to investigate the situation of commercial medical insurance in certain groups of people in China. He firstly chooses the teachers and clerks in universities and colleges as the object of study. These people are often covered by public medical care; however, the public medical care is often not enough for them due to the lack of finance, and thus many teachers choose various kinds of commercial medical insurance as a complementary insurance. This situation makes them as a good study object of investigation.

In conventional social security records, the information about the people and the health care is quite plain. Usually only the disease name, the fee and the security id are recorded. However, for a deep investigation, the researcher especially wants to know the career environment or, more specifically, the research background of the teachers in universities would affect their health and how they would choose the commercial medical insurance accordingly. The information required to accomplish such an investigation is scarcely stored in conventional social security database. Therefore, he needs to explore various kinds of information about the teachers manually. Fortunately, most of information involved can be found from the Web. For example, he can find the education and research background of the teachers in their homepages, the research and the health information from their blogs. He can also find the diseases and therapies from the Web, and all kinds of medical insurance

information from the Websites of insurance companies. Now the problem is that it is a big burden for him to gather, sort and analyze the information from the Web since he is not an expert in computer engineering. Therefore, he needs help from a technician in the computer field, or he can resort to i-SSIS to find the information.

I-SSIS is an information system which gathers information related to social security from the Web and provides typical information services to end users in the social security field. Like a search engine, i-SSIS uses the data crawled from the Web, and then analyze it to extract social security information, human resource information, etc. It deploys certain data model to organize the extracted information and uses a database system to store and query the information. Therefore, the social security data in the Web can be conveniently managed and utilized under i-SSIS.

4. Architecture of I-SSIS

I-SSIS is a database system which gathers, organizes and manages the social security data from the Web. The database system is built upon a database management system INM-DBMS as its backbone and uses certain auxiliary tools to provide various functions. The reason we chose INM-DBMS to manage social security data is that INM-DBMS adopts a novel data model named Information Network Model (INM) [9,11] which can easily associate the data about an entity and thus is appropriate and convenient to describe and manage the data of people gathered from the Web. In this section, we firstly introduce utilizing the basic features of INM to model social security data and then describe the architecture of i-SSIS.

4.1. Modelling Social Security Data with INM

INM is an object-role model which can expressively specify the attributes, roles and relationships of entities in the real world. It supports role-relationship class and class inheritance which can effectively present the complex networked semantics of entities. It can present the information of real-world entities by associating each of them to a single object with a unique oid.

In INM, an object is represented as a tree. The root of the tree is a list of object names associated with the object and each sub-tree corresponds to a property. An INM instance database is a set of classified objects.

Figure 1 illustrates a sample of INM instances which contains information about the university WHU, the course ADB, and several people Bob, Amy, Ada, Ann. The object WHU has two relationship hierarchies with roots Faculty and Student. The relationship Faculty is specialized into Prof with value Bob and Lecturer with value Amy. The relationship Student is specialized into UnderGrad with value Ada and GradStudent which is further specialized into M.Sc with value Ann and PhD with value Amy.

I-SSIS is built upon an INM instance database which is designed deliberately for social security data management. In i-SSIS the entities can be classified as 4 categories of objects: Person, Role, Organization and Insurance. A Person object represents a person entity in real world, including the attributes which is concerned in the social security such as birthday, identity (or social security) number, living place, etc. A Role object represents a role which a person acts in certain circumstance and an Organization object represents an organization, e.g., a company, an institution or an association, in which a Person

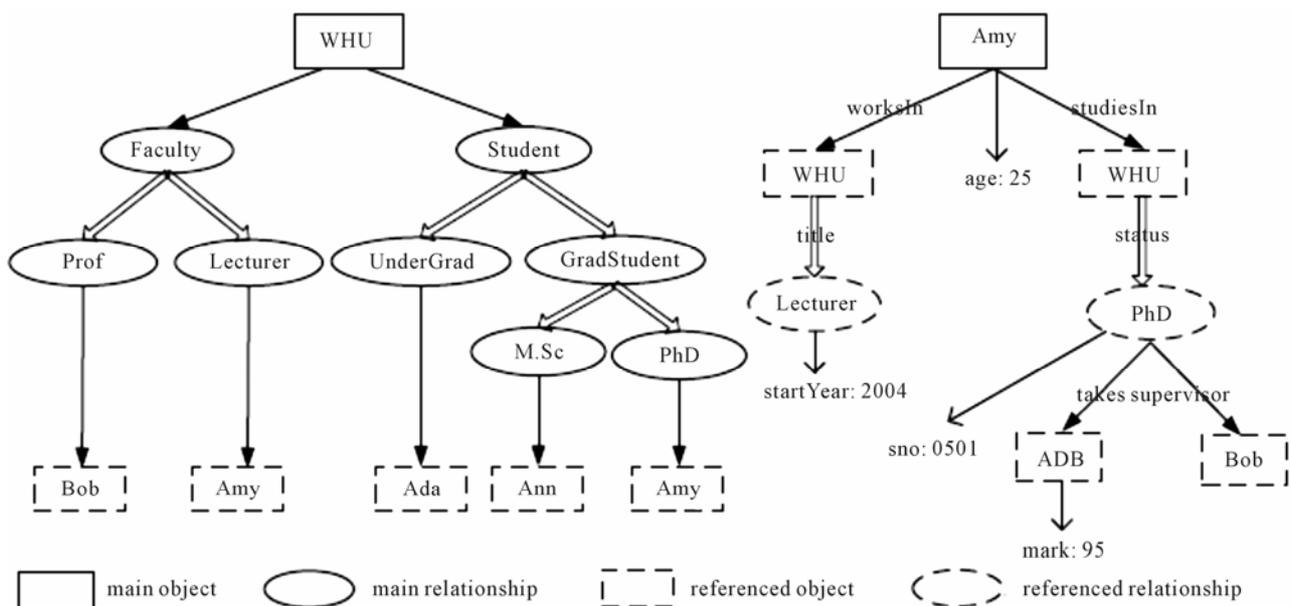


Figure 1. A sample of INM instances.

plays a certain role. For example, a person *Bob* is a faculty member in a university; meanwhile he is also an athlete in his spare time and belongs to a sport association. As INM indicates, the role can have related attributes which the related person pertains. Therefore, the objects *Person*, *Role* and *Organization* can work together to establish a full background of the people to be concerned. An *Insurance* object represents a kind of social security insurance such as pension or healthcare.

4.2. Architecture of I-SSIS

As a system to gather and manage social security data, i-SSIS has a 3-layer architecture to undertake the functionalities of gathering, managing and serving in each layer respectively, as **Figure 2** shows.

Under i-SSIS is the raw data in the Web, which is crawled and processed by the modules in the Data Collection Layer. The crawlers in i-SSIS fetch the pages from the Web sites to gather data involving human resource and social security information. The crawlers are embedded with some inner analyzers to filter the unnecessary information during crawling, which means that only the documents involving social security information such as the person, the organization, the role and the insurance are collected. Then the raw data would be processed through three procedures to become the entities managed by the database system. Firstly, the information of i-SSIS entities, e.g., people and organizations, is extracted with information preprocessing tools such as natural language processing tools. In this procedure, Web documents are initially summarized with a statistics tool to find its theme and then are processed by information

extraction tool to get the rough information about i-SSIS entities. For example, when the blog of a person Bob is processed, the biography, the affiliation, the occupation and the important social relations are extracted to construct his basic information. Secondly, various information of each entity is identified and condensed to generate entity objects in i-SSIS, as the data collection result of entities. As previously mentioned, in INM the entity is presented as a single object. However, it is common in information extraction that a single entity, e.g., a person, is described in different document fragments from different aspects. Therefore, in this layer, data mining tools and other analyzing tools are used to combine the information about distinct entities and formulate it into a predefined schema of i-SSIS entities. Thirdly, the entity and relationship data collected from the Web are integrated with the conventional social security data. The latter is collected from social security databases through common interfaces or extraction tools of deep Web. By combining the data on entities, the integrated data warehouse can provide a consistent and comprehensive description to the entities involved in social security. After the three procedures, the Web-oriented social security data is finally gathered and provided to the Data Management layer.

Processing the raw data to get social security related information is fundamental in building a practical i-SSIS. The research and implementation of the tools in this stage is being undertaken, and we have already built a prototype which can semi-automatically gather and process the information of the people in education organizations such as universities because of the abundance of Web documents about the people and organizations in this field. The crawlers gather the documents from the universities, the research institutes, the homepages and the sample social security databases. With the raw data, the i-SSIS entities, *i.e.*, the people, the organizations, the social security records, etc., are recognized and fused to reflect the entities in real world. However, since the information extraction and analyzing tools are quite difficult to be customized to fit for the practical work. Lots of work has to be undertaken manually. We are improving the tools to make data collection more efficient.

In the Data Management Layer, i-SSIS directly deploys INM-DBMS to store, manage and query the social security data provided by the under layer. The information in i-SSIS usually lies in the four categories of INM objects as described in the last section, and their storage, query and index are managed by INM-DBMS. To make the database efficient to social security data management, the INM-DBMS module in i-SSIS is especially customized to speed up the common data manipulations. On one hand, the data storage is optimized to cluster the data about entities if possible, because there are often query

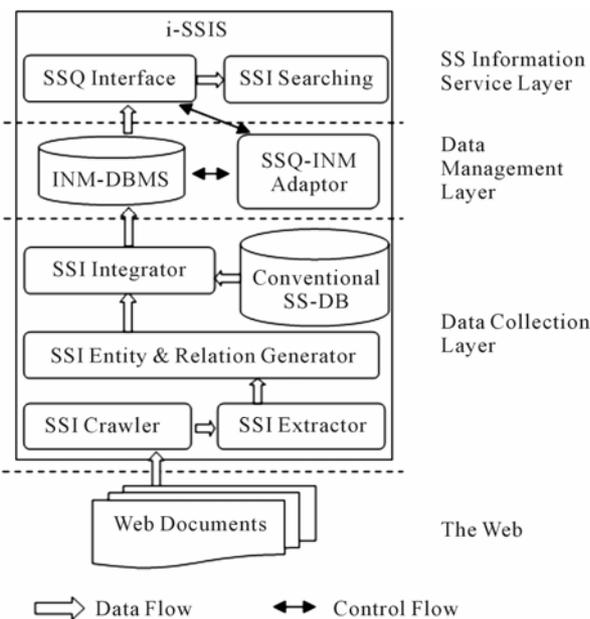


Figure 2. Architecture of i-SSIS.

requests to find associated attributes on people or organizations in practice to fetch related social insurance data. On the other hand, temporal data manipulation, a common feature of human resource information management, is also allowed here and the storage and the index are optimized to make the temporal queries to be processed more convenient and quickly.

In the Social Security Information Service Layer, a query interface named SSQ, standing for Social Security Query, is provided as the major component of the Human-Computer interface of i-SSIS. SSQ is much more than a simple application of INM-DBMS's query [10] interface because it uses a temporal query language named HRQL as the intermediate query language for the data management layer. HRQL can easily present query requirements on temporal information of person entities. This query language is an extension of our previous study [12], and the feature of entity—object correspondence of INM-DBMS is properly utilized in processing the HRQL. The SSQ interface transforms the user's query requests to the typical queries implemented in HRQL, and then forward the parsed queries to the under layers. Besides the query service, a novel social security information searching service is proposed in establishing i-SSIS and discussed in brief in the next section.

5. Searching Service of I-SSIS

Social security information service is a special part of social security data management because it provides users the approaches to utilize the social security data in practice. Conventionally, the information service is only the common query service provided by the database system. However, in the Internet era, everything is associated with searching. Therefore, it is important and challenging to build a searching service especially for social security information to make the work more efficient and convenient.

In i-SSIS, we propose and design a social security information searching service under the assistance of the Pluto searching engine [13] which is supported by INM-DBMS to undertake the INM data searching. The searching service deploys a structured data searching model. The model specifies the Steiner-trees containing all the keywords as the searching results. It ranks the results according to the compactness of the tree, the authority of the nodes and the redundancy of information. This searching model can enable the users not familiar with the INM model to utilize the social security information in i-SSIS without being aware of the networked structure of underlying entities and relationships. It can not only increase the quality and precision of searching results but also present users useful semantic relationships and context information of entities.

The searching service utilizes a heuristic searching approach based on pruning matching nodes. This approach can enhance the searching efficiency by pruning the top-k matching nodes which have minor possibilities. Based on this approach, the underlying INM-DBMS in i-SSIS deploys a special index system to support pruning matching nodes. The index system collects the neighborhood information of the nodes and utilizes it to calculate latent matching nodes. The undergoing prototype shows that the searching service can provide users a novel and efficient experience of utilizing social security data distinct from the conventional one.

6. Conclusions

Social security data management is an important topic both in application of information management and social security management. In the Web 2.0 era, more and more human information and healthcare information is released to the Internet through various approaches. This abundance makes managing social security data go beyond managing conventional social security database records. How to organize the conventional records together with the related information gathered from the Web is an interesting problem to solve to provide more convenient and powerful social security information service.

In this paper, we introduce our initial work in building a Web-oriented social security information system named i-SSIS. I-SSIS is a special database management system INM-DBMS which is used to describe and manage the real world entities using a data model named INM. Under the support of INM-DBMS and other preparation and query tools, i-SSIS can efficiently manage the social security information and provide useful service for common or specific purposes. Up to now, i-SSIS is still in its initial design phase, and we are working at building efficient information extraction tools to gather and analyze the social-security related information from Web resources. In the next step we would integrate the social security extraction system with i-SSIS and to provide a prototype for searching and utilizing the social security information.

7. Acknowledgements

This paper is partially supported by the Key Research Funds of Hubei Small and Medium-sized Enterprise Research Center under contract No.WH2011001 and the National Social Science Foundation of China under contract No.09CZZ032.

REFERENCES

- [1] Department of Health of the United States, "Second Review of a New Data Management System for the Social

- Security Administration,” National Academies Publication, Washington DC, 1979.
- [2] P. A. Diamond, “A Framework for Social Security Analysis,” *Journal of Public Economics*, Vol. 8, No. 3, 1977, pp. 275-298. [doi:10.1016/0047-2727\(77\)90002-0](https://doi.org/10.1016/0047-2727(77)90002-0)
- [3] S. Wu, Y. Zhao, H. Zhang, C. Zhang and L. Cao, “Debt Detection in Social Security by Adaptive Sequence Classification,” *Proceedings of the 3rd International Conference on Knowledge Science, Engineering and Management*, Vienna, 25-27 November 2009, pp. 192-203.
- [4] H. Zhang, Y. Zhao, L. Cao, C. Zhang and H. Bohlscheid, “Customer Activity Sequence Classification for Debt Prevention in Social Security,” *Journal of Computer Science and Technology*, Vol. 24, No. 6, 2009, pp. 1000-1009. [doi:10.1007/s11390-009-9288-2](https://doi.org/10.1007/s11390-009-9288-2)
- [5] P. Lucas, “Bayesian Analysis, Pattern Analysis, and Data Mining in Health Care,” *Current Opinion in Critical Care*, Vol. 10, No. 5, 2004, pp. 399-403. [doi:10.1097/01.ccx.0000141546.74590.d6](https://doi.org/10.1097/01.ccx.0000141546.74590.d6)
- [6] L. O. Gostin, J. T. Brezina, M. Powers, R. Kozloff, R. Faden and D. D. Steinauer, “Privacy and Security of Personal Information in a New Health Care System,” *The Journal of American Medical Association*, Vol. 270, No. 20, 1993, pp. 2487-2493. [doi:10.1001/jama.1993.03510200093038](https://doi.org/10.1001/jama.1993.03510200093038)
- [7] J. A. Lyman, K. Scully and J. H. Harrison, “The Development of Health Care Data Warehouses to Support Data Mining,” *Clinics in Laboratory Medicine*, Vol. 28, No. 1, 2008, pp. 55-71. doi.org/10.1016/j.cll.2007.10.003
- [8] S. M. Heathfield, “Human Resources Information System (HRIS)-HRIS Definition,” Technical Report, 2011. <http://www.about.com>
- [9] M. Liu and J. Hu, “Information Networking Model,” *Proceedings of 28th International Conference on Conceptual Modeling (ER 2009)*, Gramado, 9-12 November 2009, pp. 131-144.
- [10] J. Hu, Q. Fu and M. Liu, “Query Processing in INM Database System,” *Proceedings of 11th International Conference on Web Age Information Management (WAIM 2010)*, Chengdu, 15-17 July 2010, pp. 525-536
- [11] J. Hu and M. Liu, “Modeling Context-Dependent Information,” *Proceedings of 18th ACM Conference on Information and Knowledge Management (CIKM 2009)*, Hong Kong, 2-6 November 2009, pp. 1669-1672.
- [12] L. Luo, H. Yang and X. Li, “Towards Human Resource Information Query on Temporal XML,” *Proceedings of 3rd International Conference on Internet Technology and Applications (ITAP 2012)*, Wuhan, 18-20 August 2012.
- [13] The Pluto Searching Engine. <http://pluto.whu.edu.cn>