

A Reliable Organization Information Managing System based on Information Verification

ChulSu Lim*, Wongoo Lee, Yoon Young Joon and Kangryul Shon

Korea Institute of Science and Technology Information, South Korea; cslim@kisti.re.kr

Abstract

Organization information plays an important role in information systems, which generally modeled as an entity in the database. In general, the organization could be an actor or a target of the action in a business modeling. But the name of an organization may have many variants when used as an input. Those diverse usages of the organization names cause many problems in many practical applications due to lack of identification. To deal with identification problem of the organization, we develop a reliable organization information managing system based on information verification. To verify the names of the organization we utilize business identification number. This number is an analogy to a social security number to identify a citizen. Also, we constructed heterogeneous name management service to merge the variant name expressions of an organization. With the verified organization information and variant name management, the organization information could service critical business applications including contract, sanction, and information search.

Keywords: Credit Information, Heterogeneous Name, Organization Information, Verification

1. Introduction

In a database, agents are modeled as entities and actions between the agents are modeled as relations. Typical agents are person, object, and organization. So the organization information plays an important role in information systems. In general, the organization could be an actor or a target of the action in a business modeling. But the name of an organization may have many variants when used as an input. Those diverse usages of the organization names cause many problems in many practical applications due to lack of identification. This paper surveys on the usage the organization in a national R&D information management service in Korea: NTIS (National science and Technology Information Service)^{1,3}. The service provides overall Korean national R&D information. This service provides government-funded R&D information on topics as programs, projects, human resources, equipment/facilities, and outcomes^{1,2}.

The system focuses on reducing errors to enhance quality of the data^{2,4,5}. To improve data quality, this system not only consider entity relations, dependencies at the design stage but also consider data constraints of an element or concordance among related entities. These data constraints are gathered and managed as business rules for the system⁴⁻⁷. Among the efforts, information regarding organization plays an import role. Using the derived usage of organization information, we analyze needs for the integrated management of the information. To solve the problems in the system, we introduce heterogeneous name resolving, verification of identification number, and credit information support.

2. Organization Information

Organization information is one of major information items almost all system should incorporate in. And the information is very useful to data mining from text.

*Author for correspondence

So, the recognition of named entity is a fundamental task for natural language processing. The named entities include person, organization and location. And the term Named Entity was introduced in the sixth message understanding conference^{8,14}. Learning methods for the problem of named entity classification. There are many efforts to recognize and classify the named entity as a fundamental task for further comprehensive language processing⁹. Figure 1 shows an information scheme for information about organizations in NTIS service. Not only registered name, but also many heterogeneous names could be used, which may result in failures on identify a specific organization. Acronym, nickname and unofficial name could be the source of the multiple names. These multiple names are the sources of problems in information systems. First, users could not know which name should be used. When the user happened to find a name for the company, the user might miss many other variant names of the organization. Another problem is related with statistics. Information could be scattered into many organizations, which are actually the same company. The result statistics could be misleading. These problems could be solved when we could use fixed organization lists. Instead of using free text input for the names of the organizations, the system could provide lists of the names the user could choose. But for a full scale service system it is nearly impossible to manage the list of the organizations. Even though a system could afford to have such elaborate organization list management, there is a need to handle change management of the organization names. Additional information which is needed to manage projects could be divided into contact information, general information including current status of the organization, and credit information. The contact information is needed to send e-mail or packages to the organization. The general information includes total sales, pure profit and number of employees of the organization. The credit information shows the soundness of the organization. If the contracted organization shutdown their business with various reasons, the funding agency should reconsider the current contract. The agency could use this information to withdraw their investment to the organization. Figure 2 shows the usage of organization information in R&D processes. It is important to establish solid business process model through the understanding of target application after a detailed analysis on the related business processes^{10,11}. Among the processes, proposal registration, contract,

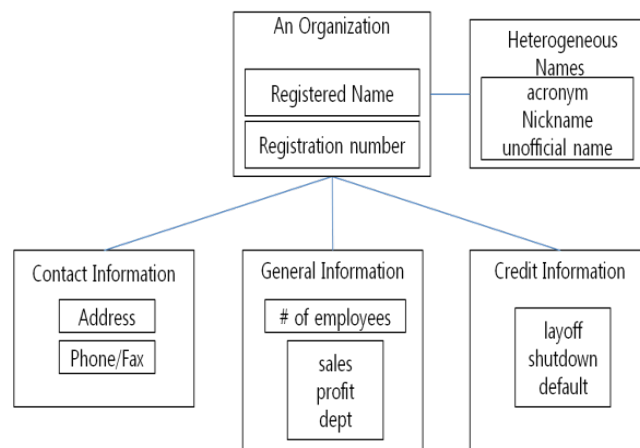


Figure 1. A Scheme of organization information

looking up information and output registration are the most important steps which require organization. In addition, during performing research step, managing institutes need to monitor the change of credit information such as shutdown or layoff of the company with which they contracted. The first row depicts processes related with the funding agency. And the second row lists processes conducted by the researchers. As shown in this figure, many processes are related to the organization information. And the credit information of a contracted organization is valid from the initial evaluation process to the final performance assess evaluation process.

Table 1 shows which information type could be utilized for each project managing processes. Registered identification number verification service is widely used throughout the processes. Heterogeneous names information is used for registering information to the system. Each service relies on the information collected in an organization information database. The information is a collection of representative name, organization status including 'close temporarily', 'shutdown businesses' and 'default on monetary debt'.

3. A Reliable Organization Information Managing System

To support the needs of the system, we analyzed needed functions and system supports for the processes⁸⁻¹³. Figure 3 depicts the relations among them. The management system incorporates three functions as APIs (Application Programming Interface). The heterogeneous name management API provides representative names when given

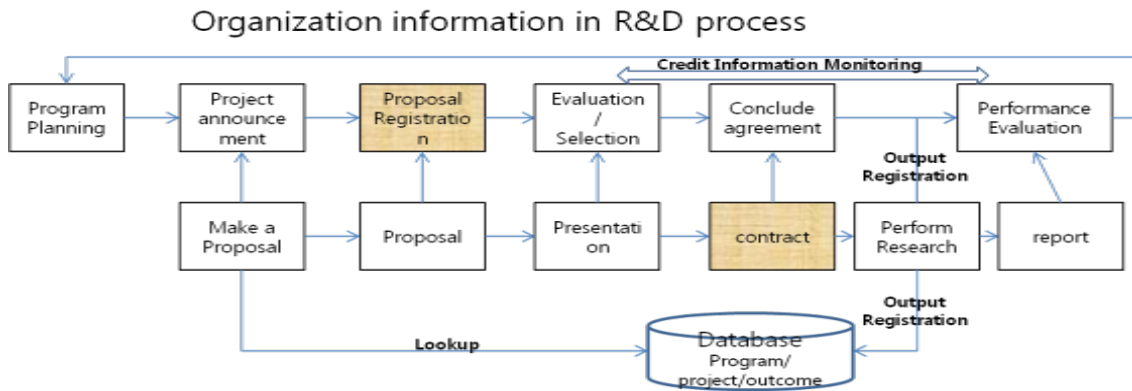


Figure 2. Usage of organization information in R&D processes.

Table 1. Usage of organization information in the r&d management

	Planning	Proposal Accept	Project Agreement	Conducting Project	Outcome Register	Performance Evaluation and Sanctioning
Service	- Heterogeneous Names - Registered ID Verification	- Registered ID Verification	- Registered ID Verification - Credit Information Lookup	- Registered ID Verification - Credit Information Lookup	- Heterogeneous Names	- Heterogeneous Names - Registered ID Verification
Information	- Representative Name - Status (Close temporarily/shutdown)	- Status (Close temporarily/shutdown)	- Status (Close temporarily/shutdown) - Default on monetary debt	- Status (Close temporarily/shutdown) - Default on monetary debt	- Representative Name	- Representative Name - Status (Close temporarily/shutdown)

heterogeneous name as an input. The registration identification number verification API validates the soundness of the name and identification number of an organization. The credit information management API provides information about the status of an organization like shutdown, layoff, and bankruptcy.

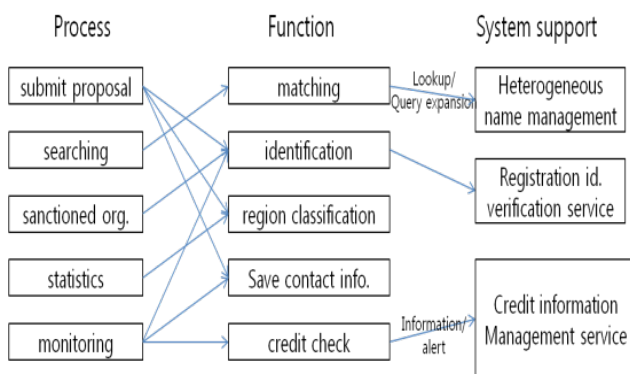


Figure 3. System supports to serve the information usage.

The heterogeneous name management encourages representative names of an organization instead of nicknames which could be a source of errors in searching functions. With the help of verification on the identification number for the organization, we could identify a unique organization instead of names, which causes many other problems for the processing of the information for statistical analysis. Figure 4 shows the relation between representative name and heterogeneous names of an organization. The representative name tends to be an official name. But in some cases the official name and major brand name could be different. In that case, the representative name could be carefully chosen based on documented rules. Thing could be complicated when the organization is big enough to have many branches and affiliated companies. In that case, rules must be adjusted to support information needs of related applications. The example in the figure shows a famous university which has many nicknames and many companies.

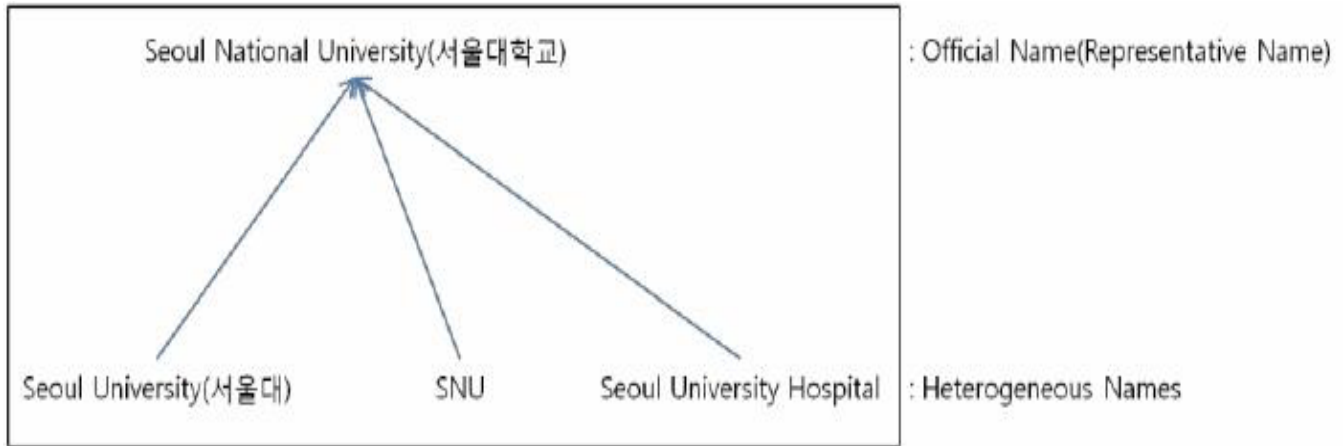


Figure 4. A representative name and heterogeneous names.

‘Seoul University’ and ‘SNU’ are nicknames and ‘Seoul University Hospital’ is a name of the subsidiary organization. A specific subsidiary organization could be selected as a representative name when the company should act as a unique role in an application service.

of an organization require a management system which needs to address the following issues: defining guidelines for a representative name methods to review and confirm the representative name

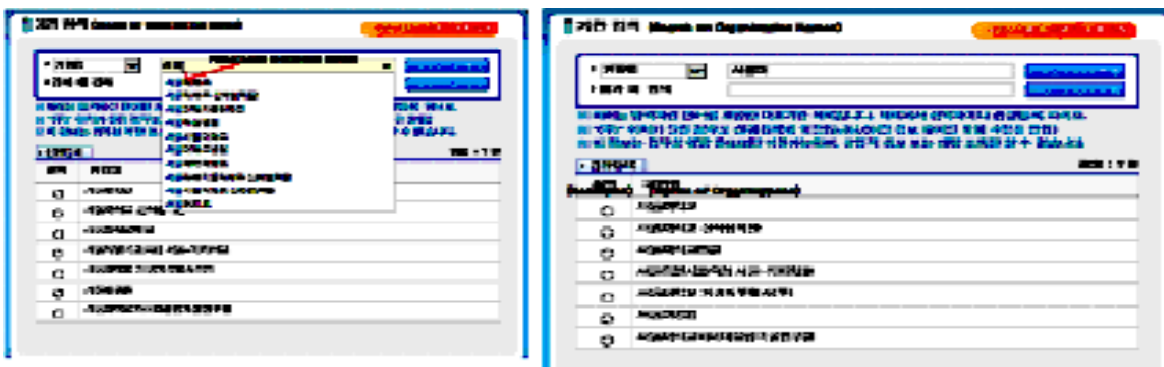


Figure 5. Search on organization names with representative name suggestion service.

The representative name could be chosen as an input using heterogeneous names related to the organization as in Figure 5. Often the case, users don’t remember canonical representative names for an organization. Instead they are accustomed to use their own names. The user interface is designed to show the candidate lists of representative names with the user’s partial input. By promoting to use the representative names, the names of the organization are clearer for later use in subsequent applications on the system. Using the representative names in the data gathering stage could prevent possible inaccuracy or errors. The benefits of using the representative names

keeping track of reference and change of the organization providing information for searching with heterogeneous names

Figure 6 shows a user interface to verify registered Identification Number (ID) of an organization. After filling up the name and registered ID of the organization, the user could verify the validity of the information by selecting the ‘Verify’ button. The validity is approved by looking up a database from the tax office. Also, users could use a certificate of authentication for the organization.

>사업자명 한국과학기술정보연구원 >사업자등록번호 205 - 82 - 04053

(Name) Registered ID No. Verify Verify with a authentication

기업실명확인 법인공인인증서 확인

Figure 6. Registered identification number and name verification.

사업자번호 Registered ID No.	2058204043	NTIS인증키 Authentication key	103%Qb0V8xW0UJ~2R1JA==
요청사업자명 Name	한국과학기술정보연구원	결과메세지 result message	정상처리 되었습니다.
요청사업자명 verified name	한국과학기술정보연구원	응답코드 response code	9002
등록상태코드 status code	G	유효인증서 status change date	기업정보관리부등록일
업종코드 M70129	1141710002397	대표자명 한선화	details
법인번호 99	기업유형코드 1	사업자명 한선화	100525169
총자산(전월) 139530192	자기자본(전월) 1500	부채(전월) 39005023	financial information
납입자본금(전월) 119978510	영업이익(전월) -909481	결산기준일 20131231	20090501
당기순이익(전월) 339	종업원수기준일 20090501	Address	
(구)주소우편번호 305806	(구)주소 대전 유성구 여은동	도로명주소 305806	도로명주소 대전 유성구 여은동
(구)상세주소 52-11번지	도로명주소 305806	도로명주소 305806	도로명주소 305806
도로명주소우편번호 305806	도로명주소 305806	도로명주소 305806	도로명주소 305806
도로명상세주소 245 (여은동)	도로명주소 305806	도로명주소 305806	도로명주소 305806
응답코드 (Response code)			
9002	일치(법인번호 요청안함, 업체명 O)	9010	불일치(법인번호 요청안함, 업체명 X)
9011	사업자 없음(법인번호 요청안함, 업체명 KED 없음)	9012	사업자 없음(법인번호 KED없음, 업체명 KED 없음)
기업유형코드			
1	법인	2	개인
유효인증서코드			
G	정상	C	취소
O	폐업	K	국세청 조화실파 & KED미등록
N	미등록 사업자번호		

Figure 7. Response of the registered Identification and name verification service.

사업자번호 (Registered Id No.)	2058204043	요청사업자명 (Name of an organization)	한국과학기술정보연구원
응답코드 (Response code)	9024	메세지 (Message)	유효한 인증서가 아닙니다.
NTIS인증키 (Verification Key)	ak7KsD4B376IVP70M7mPBQ==		
응답코드(Response code list)			
9021	공인인증서-인증완료	9022	공인인증서-사업자번호 불일치
9023	공인인증서 유효성 검증실패(Client Side)	9024	공인인증서 유효성 검증실패

Figure 8. Response of the registered Identification and name verification service.

The response message of the verification not only approves the validity but also provide detail information needed for the management of R&D information. As in Figure 7, the response code indicates the validity of the information. And the additional information include status of the organization, authentication key, financial status, address and detail classification codes for the organization. The authentication key could be used to validate the message itself later. The verification using a certificate of authentication for the organization utilizes stored information in the certificate itself. So, the response message could tell the soundness of the identification number of the organization as shown in Figure 8.

Another type of information needed for the management of project is credit information of the participant organizations of the project. When registering a new project proposal, the manager needs to check the soundness of the organization. After a contract for a project the status of the organizations involved in the project needs to be checked. For an organization which closed their business in the period of the contract, the contract needs to be retracted. It may be also related to a funding expenditure management system. Figure 9 shows an example list of credit information for organizations. Users could check the recent changes of status of credit information of organizations. Also, the users could get alarm messages for some events they need to reconsider the contracts.

Not only closing temporarily or shutdown of their business, but also failure to pay off their debts raises an event.

we develop a reliable organization information managing system based on information verification. To verify the names of the organization we utilize business

No.	Identification No.	Organization Name	Request Date	Confirmed	Recent Shutdown Records	Default	Alarm
27	110-15-21247	유비넷비	2014-07-11 10:25:22	Y	폐업자 (2015-02-15)	Y	OFF
26	121-81-38248	삼진엔테크(주)	2014-07-11 10:25:22	Y	정상-부가가치세 일반과세자 (2015-02-08)	N	ALARM
25	209-81-27944	한메덱스 주식회사	2014-07-11 10:25:22	Y	정상-부가가치세 일반과세자 (2015-02-08)	N	ALARM
24	215-81-52337	열림디비산업(주)	2014-07-11 10:25:22	Y	폐업자 (2015-02-15)	N	ALARM

Figure 9. List of credit information for organizations.

Identification No.		110-15-21247		Name of the Organization		유비넷비	
No.	Date	Date of Event	Seq. No.	Name of the Bank		Registered Reason	
8	2013-11-20	2015-02-24	0000002	기술신용보증기금 종로기술평가센터		신용보증대지급금	
7	2013-08-26	2015-02-24	0000001	기업은행 홍제동(지)		신용카드연체	
6	2013-11-20	2014-12-21	0000003	기술신용보증기금 종로기술평가센터		신용보증대지급금	
5	2013-12-30	2014-12-21	0000002	신용보증기금 마포(지)		신용보증대지급금	
4	2013-08-26	2014-12-21	0000001	기업은행 홍제동(지)		신용카드연체	
3	2013-11-20	2014-12-19	0000003	기술신용보증기금 종로기술평가센터		신용보증대지급금	
2	2013-12-30	2014-12-19	0000002	신용보증기금 마포(지)		신용보증대지급금	
1	2013-08-26	2014-12-19	0000001	기업은행 홍제동(지)		신용카드연체	

Figure 10. The credit information about a contracted organization.

As shown in Figure 10, the project managing institutes could check the status of their contracted project participating organizations with the detailed credit information. The credit information is provided by a federation of banks. The heterogeneous names of a company name promote utilizing official name by suggesting representative official name when users enter the name of the organization. And this information broadens coverage of search result by expanding search keywords. This is similar to search query expansion using synonyms of the keywords. The verification of the identification number of the company promotes utilizing quality data. Identification of entities in a system is one of the fundamental issues for information systems. The credit information of companies helps informed decision making in the processes of project management.

4. Conclusion

To deal with identification problem of the organization,

identification number. This number is an analogy to a social security number to identify a citizen. Also, we constructed heterogeneous name management service to merge the variant name expressions of an organization. With the verified organization information and variant name management, the organization information could service critical business applications including contract, sanction, and information search. The heterogeneous name management encourages representative names of an organization instead of nicknames which could be a source of errors in searching functions. Using the representative names in the data gathering stage could prevents possible inaccuracy or errors. The heterogeneous names of a company name promote utilizing official name by suggesting representative official name when users enter the name of the organization. And this information broadens coverage of search result by expanding search keywords. With the help of utilizing registered identification number of the organization, quality of data about the organization could be enhanced. Finally, the agile application of the credit information to monitor the organization

under a contract supports upgraded management of the research and development projects for the managing institutes.

5. Acknowledgement

This research was supported by the Sharing and Diffusion of National R&D Outcome funded by the Korea Institute of Science and Technology Information.

6. References

1. Song CH, Sul SS. A study on the problems of current national standard classification of science and technology for national science and technology information system. *Journal of Korea Technology Innovation*. 2006; 496–513.
2. Lim C, Kim J-M, Yoon Y-J, Shon K-R, Kim J-S. Data content constraint management for national R&D data quality improvement. *International Conference on Convergence Contents*; 2011. p. 17–8.
3. Available from: <http://www.ntis.go.kr>
4. Zimbrão G, Mir R, De Souza JM, Mario H, Neto FP. Enforcement of business rules in relational databases using constraints. *SBBB*; 2003. p. 129–41.
5. Lim CS, Shon KR, Kim TH, Han SW, Lee WG, Kim JM. Improvement of R&D information management to support convergence research. *International Conference on Convergence Technology*; 2013.
6. Ceri S, Widom J. Deriving production rules for constraint maintenance. *Proceedings of 16th International Conference on VLDB*; 1990 Aug 13-16; Brisbane, Australia: 566–77.
7. Ceri S, Cochrane R, Widom J. Practical applications of triggers and constraints: success and lingering issues. *Proceedings of 26th International Conference on VLDB*; Sep 10-14; Cairo, Egypt: 2000. p. 254–62.
8. Mansouri A, Affendey LS, Mamat A. Named entity recognition approaches. *International Journal of Computer Science and Network Security*. 2008; 8(2):339–44.
9. Collins M, Singer Y. Unsupervised models for named entity classification. *Proceedings of the Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora*; 1999. p. 100–10.
10. Georgakopoulos D, Hornick M, Sheth A. An overview of workflow management: from process modeling to workflow automation infrastructure. *Distributed and Parallel Databases*; 1995.
11. Abreu FBE, Freitas JM. Definition and validation of metrics for ITSM process models. *Seventh International Conference on the Quality of Information and Communications Technology*; 2010.
12. ISO/IEC 20000-1: Information technology - Service management - Part 1: Specification. Standard, International Organization for Standardization (ISO/IEC); 2005.
13. Gruhn V, Laue R. Approaches for business process model complexity metrics. In: Abramowicz W, Mayr HC, editors. *Technologies for Business Information Systems*. Springer; 2007. p. 13–24. ISBN:978-1-4020-5633-8.
14. Lee S-H, Lee D-W. A case study in Japanese and prospect of cloud computing service in convergence age. *Journal of the Korea Convergence Society*. 2015; 6(1):17–22.