

DEVELOPMENT PROCESS OF INFORMATION FLOW RETRIEVAL SYSTEM FOR LARGE-SCALE CONSTRUCTION PROJECTS

Jinho Shin¹, Hyun-soo Lee², Moonseo Park³, Jung-ho Yu⁴ and Jungseok Kim⁵

¹ Ph.D. Candidate, Seoul National University, Seoul, Korea

² Professor, Seoul National University, Seoul, Korea

³ Associate Professor, Seoul National University, Seoul, Korea

⁴ Assistant Professor, Kwangwoon University, Seoul, Korea

⁵ Ms. Student, Seoul National University, Seoul, Korea

Correspond to beyond@snu.ac.kr

ABSTRACT: Players of construction projects proceed with each work process by information gathering, modification and communication. Due to the complex and long-span lifecycle projects increased, it became more important to grasp this mechanism for the successful project performance in construction project. Hence, most project information management systems or knowledge management systems equip information retrieval system. There are two logic to infer the meaning of retrieval target; inductive reasoning and deductive reasoning. The former is based on metadata explaining the target and the later is based on relation between data. To infer the information flow, it is necessary to define the correlation between players and work processes. However, most established information retrieval systems are based on index search system and it is not focused on correlation between data but data itself. Thus, this research aims to research on process of information flow retrieval system for large-scale construction projects.

Keywords: Information Retrieval; Relation-Based Database Modeling; Business Process Modeling; Information Flow Analysis

1. INTRODUCTION

In construction projects, various players progress a work process during producing information and transferring it with another players. If players produce insufficient information or do not transfer in time, it becomes the reason of delay or decrease the performance quality even though it has no direct relationship. Therefore, players should comprehend overall information transference work; such as what I should produce in this work, when I should transfer and whom I should transfer.

Most problems which are caused on the information transference do not come to the surface right after the beginning of problem causes. And the characteristics of construction projects; such as diversity, complexity and various site circumstances hinder players from storage of information [10]. Especially, the large-scale construction project team is organized as the complex co-work system. Results from this team composition, the main player to generation necessary information in the early stage is usually different organization with the player to complete the final output at the last stage. Furthermore, the team members usually belong to different firms or organization and work together for one by one project. Each player have own information transference system and I arouses inconsistent system of information transference. For these reasons, players are hard to analyze the information flow and use for other projects. Thus, this research aims to

develop the information retrieval system for comprehension of information transference in large-scale construction projects.

It should be also considered of proper target business process modeling and database modeling while planning the information system. Therefore, we clarified the keynotes of information retrieval system for the large-scale construction projects at the first stage. Next, we reviewed the established information retrieval technology and studied the proper information retrieval technology. And then, we set up the deduction algorithm for the system and analyzed the each development process.

2. THE CHARACTERISTIC OF LARGE-SCALE CONSTRUCTION PROJECTS

The concept of large scale and complex projects which is mentioned in this paper does not mean the project which cost considerably. It involves multi projects which are compositive relationship, long-span lifecycle and diverse players.

Most large scale projects are public projects such as bridge, highway or tunnel and they are single-purpose projects. However, multi-purpose projects such as district-units projects and urban regeneration projects are also increasing recently in architecture construction projects. The characteristic of these projects are multiple purposes, including multi subordinate projects, organic

co-ordinate relationship, and powerful economic/social effectiveness.

In these projects, the procedure of administration works before contract phase such as project application or approval is more strict, more complex, and longer than general projects due to the effectiveness. In this phase, the most various players are participated in the whole project lifecycle; administrative agency, developer, engineering firms, local residents and so on. Because each players generates various information in own format, it is hard to grasp the information flow. Thus, in the case of large and complex construction projects, it is necessary to develop support tool based on business process for retrieval complex information flow mechanism between various players.

3. THE OVERVIEW OF INFORMATION RETREIVAL SYSTEM

To search the information transference mechanism, players should know about the relation between business works accurately and promptly. However, it is hard to search it using the index search system which is the most general information retrieval system. Because, the index based retrieval system includes only query tables which are consist of metadata explaining the data itself. To improve it, the relation based retrieval system has studied in several researches ([7], [8], [9]). The relation based information retrieval system enable to search the semantic relation between data, but the optimized methodology has not been addressed. One of the most crucial reason to impede the improvement of the relation based structure, the definition of data is different by purpose, perspectives and scopes. Although the scoping study is the important in the searching process of relation based retrieval system, it is not been focused relatively. However, it is necessary to clarify the appropriate methodology of the information retrieval system for the practical use. Hence, we reviewed the established search system and ducted the most proper search system for the information flow analysis.

3.1 Directory based Search

The directory based search is the earliest retrieval technology. A system manager arranges the collected data to directory under self-standard and a user tries to search for the target information based on each directory. In this system, the purpose of the system manager can be reflected directly and user can gain the handpicked data classification under the constant standard. On the other hand, the classification standard of the manager is possible to differ from the user's it, and so it occurs to impede the searching tendency of the user. Furthermore, the human manager should manage the system for himself/herself. Because of its means, it has limitation for backup and updating the amount of information which is generated geometrically.

3.2 Index based Search

It is the 1st generation technology of the robot search engine and the most general information retrieval technology until these days. Search engine extract the

data automatically including tags which correspond with input keywords. The technology enables to increase the permissible capacity of information. And, it enables to reveal the user's searching autonomy comparing to directory base search. According to increase of handling data, it becomes matter to develop the algorithm for cut the searching time.

The retrieval process of index based search is mostly different by the size of database. In the small size database as the home network, search engine retrieves the whole database and extract data corresponding with keywords. By this means, it enables to output the result without exception. However, if the database is too huge and complex, the searching time increase and it drops practicality. For the reason, the general web based search engine uses index crawling.

The crawler in the search engine is not connected with a server directly but it crawl the data and send to the indexer for indexing the data. The indexed tag is attached to data that user can retrieve and the data stand by in the runtime indexer. Passing over this process, the search engine responds to the search keywords and results the data with the tag standing by in the runtime index.

According to the crawling and tagging algorithm, the output of index based search engines is different. And regulated extracting process is hard to reflect the purpose of information searching. To complement these weaknesses, the directory based search and the index based search are generally used in combination.

3.3 Natural Language Search

Natural language search enable to indicate the question or the sentence as the search keywords and output the results. The search results are also illustrated in natural language. The retrieval process consists of query database composition, input query analysis, result output, and user feedback. First, the search engine analyzes the composition of query. And then, it extracts the relevant query data from the pre-structured query database. The query data is recomposed to natural language and served to users. The accuracy of the system depends on the query database. To enhance the accuracy of the search result, it is used to structure the major natural language database.

The natural language search enables to retrieve the information by daily expression but, the input format is limited technically. The vague sentence decreases the accuracy of result conspicuously. Because of it, it is considered to be more acceptable to specific database. However, the usefulness of the natural language search is uncertain comparing with system building cost.

3.4 Thesaurus and Ontology

Thesaurus and Ontology is the most popular relation based search system. Thesaurus is a kind of domain concept modeling which structuralizes the semantic relationship of terminologies. This modeling defines several simple relation such as BT(Broader Term), NT(Narrower Term), RT(Related Term). Ontology enables to express various semantic relationships between the schematic factors and the semantic characteristics in

the domain. It provides the knowledge-use function through sharing, reusing and semantic deducting.

Thesaurus and ontology is relevant from the perspective of setting semantic structure of domain in detail. Because of it, thesaurus might be considered as the light ontology. Ontology is based on artificial intelligence and so, it enables to semantic deduction and various expression of semantic relationships.

However, the scope and standard of semantic relation have not been reached to an agreement and the system build costs considerably. Furthermore, there are diverse assignments such as relation set-up plan for the added information.

3.5 Information Retrieval Technology for Searching Information Flow

Table 1 shows the overview of transition of retrieval system. As mentioned earlier, the most general information retrieval system is based on index. However, it is hard to search the relation between data. Although most appropriate information technology to search the relation between data is the thesaurus or ontology, they have several critical limitations in scope and standard. Therefore, it is necessary to define the scope and standard for the information flow retrieval.

Table 1. The overview of transition of retrieval system

	Overview	System Manager	limitation
Directory based	<ul style="list-style-type: none"> ● The first age searching technology ● A system manager classifies contents by itself. 	● Human	<ul style="list-style-type: none"> ● Difficult ● Enormous data ● Impede the user's discretion
Index based	<ul style="list-style-type: none"> ● The most general methods ● Suitable for managing enormous data 	● Computer	<ul style="list-style-type: none"> ● Less systemized due to crawling in a lump
Natural Language	<ul style="list-style-type: none"> ● Sentence query 	● Computer	<ul style="list-style-type: none"> ● Limited query ● Considerable set up cost
Thesaurus and Ontology	<ul style="list-style-type: none"> ● An correlation meaning of contents based search 	● Human and Computer	<ul style="list-style-type: none"> ● Ambiguous standard of setting the semantic relation ● Considerable set up cost

4. DEDUCTION ALGORITHM

As mentioned earlier, the most proper information retrieval system among the reviewed ones is the ontology. Because, the deduction algorithm is essential to search the activity order for searching business process relation but the index search engine enable to search the data only including keywords in the data table. However, the semantic relationship increases geometrically according to increase of data, so it is unpractical to define the all of semantic relationship. Furthermore, defining whole relation is far from the decision making support and exceed information providing might cause the confusion to player. Therefore, it is unnecessary to define the whole data relationship but the selected kernel relationship which enables to grasp the information transference mechanism.

Considerable information and data enable to support understanding work more specifically. Nevertheless, in

the cooperation work, it is necessary to handle each work quickly and accurately, not to acquire whole information of corresponded work. It is easy to judge that more information enable the more accurate business work. In fact, essential information can bring work to more appropriate performance than excessive unrefined information. Because, information which has slight relevancy with work could hurt the focus of work objective and miss the proper decision moment. Thus, how to extract the kernel information from enormous information is related with successful work performance.

For the development of information transference retrieval system, it is the first step to develop the data storage for the information which is generated or used in business works. On the perspective of business process activity, the information could be classified to two kinds of data. One is the data which explain the details of work, and another is the data which explain the flow of work process. The former can be classified to metadata and some of them are included in the later. The former include data such as definition, method and matters to be attended to. The latter include data which can information flow. For instance, if "In the design phase, the design plan is produced and in the construction phase, the building work is progressed according to the design plan", two phases have information flow order relationship. Among those, former is supposed to be more valuable for decision making for information transference in the large-scale construction projects than latter.

On this perspective, the information transference mechanism consists of the target information and players. The player takes roles of the main agent of information generating as well as the main agent of information accepting. And the information takes roles of the preparation material or information for the activity as well as the outcome of the activity. The information flows transforming on the progress, the information transference between the processes and players can be defined by this flow. For example, player X associated with process A accepts the preparation material or, develops it and transfers it as outcome to player Y associated with process B. (Figure 1.). Player Y accepts this information as preparation material. If the outcome of the process A and the preparation material of process B are same, the relation of process A and B is defined as information transference relation. The relation of player X and Y is as well. If the system focus on the information flow using this deduction algorithm, it is supposed to become the more practical and useful system.



Figure 1. The deduction of the information flow

5. BUSINESS PROCESS MODELING

Nevertheless the importance of designing the business process modeling with the design of their corresponding information systems should be apparent in theory, such integrated consideration rarely have been in practice [1]. For development of the efficient retrieval system, the business process modeling associated with information system is a matter claiming prior settlement. Scholz-Reiter and Stickel described the term of business process modeling clearly, which is used to incorporate all activities relation to the transformation of knowledge about business systems into models that describe the processes performed by organizations [3]. To make accord the objective of system, it is need to investigate the target of building information management system and clarify the relation between businesses.

However, the multiplicity of possible modeling goals and objectives might render the development of such a modeling technique impossible or at least impractical. To solve the problem, the techniques should be reviewed to focus on the modeling perspectives and objectives [1].

As mentioned in section 4, this research aims to develop the information retrieval system based on the relationship business processes and players, business work. IDEF0 (Integration DEFinition 0) focuses on the flow factors including input, output, control and mechanism and it is supposed to be acceptable for the modeling of the information flow (Figure 2.).

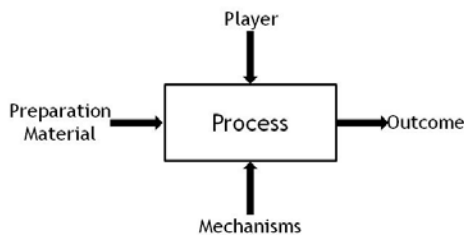


Figure 2. An adoption of IDEF0 method

6. DATABASE MODELING

The database modeling is concerned important in information management. The database modeling is

classified to two kinds by the focus. One focuses on the explanation of data itself, and the other one focuses on the explanation of the relation of data. The database model fitting for former is a hierarchical model, and for latter is a relational model [13]. The system to be developed in this research focuses on investigating the relation of the information transference. Among the relational model, ERD(Entity Relationship Diagram) is the most general methodology to illustrate the relation of activities, so we took this modeling technology to explain the information transference of business activities.

7. DEVELOPMENT PROCESS FOR THE INFORMATION FLOW RETRIEVAL SYSTEM

Information flow retrieval system consists of set up the deduction algorithm, business process modeling and database modeling (Figure 3.). In the business process modeling, the deduct factors of information transference is extracted. In the database modeling, the relation entity tables are composed of extracted factors. The retrieval system crawls the tag from this tables, and generates the index tables of information transference.

8. CONCLUSIONS

In this research, we analyzed the information transference mechanism of large-scale construction projects by input/output information and the relation of players and studied the development process for information flow retrieval system. The summary of this research is as follow.

- For the characteristic of large-scale construction project, it is important to apprehend the information flow.
- Established information retrieval systems are insufficient as they are, and it is need to scope for the target business process.

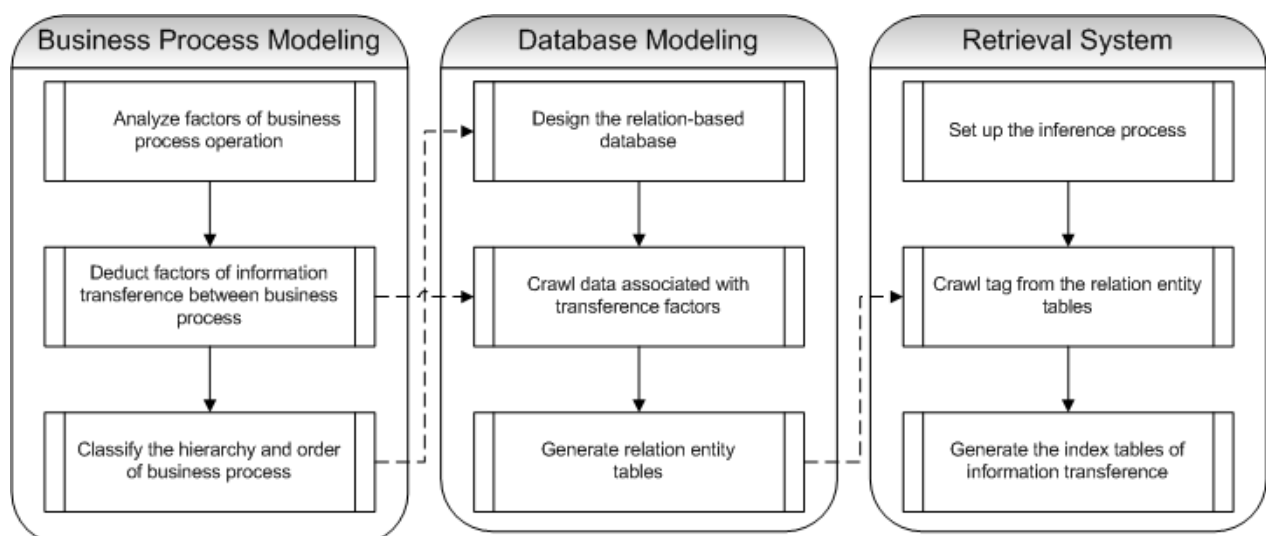


Figure 3. Development process of relation based retrieval system

- Using the relation of input and output information, the information transference between business processes can be deduced.
- The IDEF0 and ERD technology is appropriate to the development of the information flow retrieval system.

In the construction management work, the adoption of information technology is a widespread view. According to growth of complexity and size of projects, the importance of the information management has been emphasized considerably. However, it is difficult to manage the information for the diverse project players and organic co-ordination system by established information system. Especially, the information retrieval is not enough to catch the needs of information flow management. To complement this point, the semantic web such as ontology is concerned lively. However, the practical adopting planning has not suggested well because of the scope and the standard. This research addressed the alternative way to approach the relation setting. As limited to focus on the information transference, we studied the develop process for the information flow retrieval system. Although it is the uncompleted system as the relation-based retrieval system, it may be the practical suggestion for the objectives specialized system.

REFERENCES

- [1] G.M. Giaglis, A taxonomy of business process modeling and information systems modeling techniques, *The International Journal of Flexible Manufacturing Systems*, 13(2001) 209-228
- [2] W. Curtis, M.I. Kellner, J. Over, Process modeling, *Communications of the ACM*, 35(9) (1992) 75-90
- [3] B. Scholz-Reiter, E. Stickel, (eds), *Business process modeling*, Springer-Verlag, Berlin (1996)
- [4] S. Liao, Knowledge management technologies and applications-literature review from 1995 to 2002, *Expert Systems with Applications* 25 (2003) 155-164
- [5] Y. Li, Y. Wang, X. Huang, A relation-based search engine in semantic web, *IEEE Transactions on Knowledge and Data Engineering*, 19(2) (2007) 273-282
- [6] T. Mezher, M.A. Abdul-Malak, I. Ghosn, M. Ajam, Knowledge management in mechanical and industrial engineering consulting: a case study, *Journal of Management in Engineering* 21 (3) (2005) 138-147
- [7] Y.C. Lin, L.C. Wang, H.P. Tsering, Enhancing knowledge exchange through web map-based knowledge management system in construction: lessons learned in Taiwan, *Automation in Construction* 15 (6) (2006) 693-701
- [8] U. Shah, T. Finin, A. Joshi, R.S. Cost, J. Mayfield, Information retrieval on the semantic web, *The 11th International Conference on Information and Knowledge Management*, (2002) 461-468
- [9] El-Diraby, T. E., Lima, C., and Feis, B. (2005) Domain Taxonomy for Construction Concepts: Toward a Formal Ontology for Construction Knowledge, *Journal of Computing in Civil Engineering*, ASCE, Reston, VA., 19(4), 394-406.
- [10] Richard H. Clough, Glenn A. Sears, S. Keoki Sears, *Construction Project Management*, 4th ed., Wiley, New York, 2000
- [11] A. Tsalgatidou, S. Junginger, Modeling in the re-engineering process, *SIGOIS Bulletin*, 16(1) (1995) 17-24
- [12] A. Blyth, Modeling the business process to derive organizational requirements for information technology, *SIGOIS Bulletin*, 16(1) (1995) 25-33
- [13] *Data Integration Glossary*, U.S. Department of Transportation, (2001)
- [14] Maedche, A., Motik, B., Stojanovic, L., Studer, R., and Volz, R. (2003) Ontologies for Enterprise Knowledge Management. *IEEE Intelligent Systems*, 18(1), 26-33.