

# ***FRBR Implementation on a Thesis Collection in National Central Library of Taiwan: A Prototype Case Study***

陳立原

國家圖書館館員

Li-Yuan Chen

Librarian, National Central Library

Email: chein@ncl.edu.tw

陳昭珍

國立臺灣師範大學圖書資訊學研究所教授

Chao-Chen Chen

Professor, Graduate Institute of Library and Information Studies (GLIS), National Taiwan Normal University

Email: cc4073@ntnu.edu.tw

Keywords (關鍵詞) : FRBR; Thesis; Knowledge Organization

---

## **【Abstract】**

This study implements an FRBR (Functional Requirements for Bibliographic Records) model to organize master's and doctoral theses available in the National Central Library of Taiwan (NCL) and attempts to assess its feasibility. There are a few relationships in theses. So we select thesis as our object. We employ the FRBR model to analyze entities and relationships for those theses, design an algorithm, and set up a system for trial runs. On the side, we try to find other relationship for those theses. Finally, the full FRBR model of thesis is built. The study shows by its results that the FRBR model in our case is better than traditional cataloguing and indexing methods. And then, the study builds the FRBR model for the

theses. It provides the user with better and faster services in accessing the theses collection in the National Central Library of Taiwan.

## **INTRODUCTION**

The 1990 Stockholm Seminar on Bibliographic Records, sponsored by IFLA, established a core standard for a basic level record, necessary for the viability of shared cataloguing programs. After years of extensive study, IFLA in 1998 published and proposed its official FRBR data model, and it has since been met with enthusiastic responses worldwide as research and experiments have put FRBR principles into practice. Cases of FRBR application include two well-noted works: first, AustLit, a non-profit collaboration between ten Australian universities and the National Library of Australia; second, OCLC's FictionFinder in

WorldCat. To a remarkable degree, these applications, along with RDF, another work which integrates FRBR, have produced flexible cataloguing systems, similar to VisualCat's electronic cataloguing, authority controls, and indexing [1]. Moreover, related studies have been done by many devoted to the field, including application works by Hickey, O'Neil, and Toves [2], Hegna and Murtomaa [3], Bennett, Lavoie and O'Neil [4], and Berg [5].

The FRBR catalogues currently available have been produced in countries where English is the first or second language in communication; in contrast, Chinese FRBR catalogues remain to be developed. Furthermore, completed FRBR works are confined to catalogues using MARC format, while the application to dissertation collections awaits initial researches. Unique thesis features, such as knowledge-genre relationships, work-relationships, mentor/ advisor-relationships, reference-relationships, present themselves as distinctive attributes in theses collections. Therefore, this study focuses on the FRBR application to a thesis collection in the NCL and aims to explore the feasibility of the model for future wider uses.

## LITERATURE REVIEW

Since the publication of FRBR model, researchers and institutions have investigated its development and possible applications. Studies have shown two related directions taken simultaneously in the field of study. One direction is the development of infra-structural software packages in accordance with the FRBR terms of reference and their direct application to increase the speed and user-friendliness of indexing systems. The other concerns possible automated algorithmic mapping of existing cataloguing systems to an FRBR model, while the crux of this direction lies in the degree of convenience and easiness by which conversion from a traditional cataloguing system to FRBR can be achieved.

## Some Major FRBR Applications

Five major projects have been developed applying FRBR to their particular bibliographic records. First, FictionFinder from OCLC. Second, AustLit of the National Library of Australia. Third, Berg [5] applied the model to bibliographic databases and uses ACCESS indexing systems to upgrade operational power. Fourth, the University of Rochester in New York incorporated FRBR for the special archives in their library on the River Campus. Fifth, the Research Libraries Group (RLG) has taken FRBR as one of the main objectives in their RedLightGreen Project.

## Mapping of Existing Cataloguing to FRBR Model

Day [6] investigated potential mapping for automated conversion/transfer between FRBR and another metadata model, exploring algorithmic compatibility between these two metadata models as well as attempting to create a relationship between FRBR and Dublin Core attributes.

Hickey, O'Neil and Toves [2] extended the FRBR work to include additional bibliographic formats and devised an algorithmic paradigm to create possible mapping and conversion. The paradigm's logic follows:

1. Construct a key based on the normalized primary author and title.
2. If that key matches an existing set, add this record to the set.
3. If not, construct additional name/title keys based on other names and titles in the record.
4. Check each of those keys in succession. If a match is found with an existing set, add this key to that set.
5. If no matches are found, create a new set based on the original key (from MARC 1XX for the author main entry and 24X for title fields).

In their study, when constructing the keys for the algorithm, names and titles were looked up in the LC name authority file, and the established form of the name and title was used. In the process of looking up and matching, if ambiguities occurred, two methods were used to work toward resolution. First, more exclusive and stricter criteria were used to look up and match and gradually move on to other looser criteria. Second, they increased the size of collections in the dataset. Each method has its potential error rate. It has been suggested that certain manual instruments be in place to reduce errors.

Hegna and Murtomaa [3] explored devising an algorithm to map two national bibliographies, Finnish and Norwegian, coded in MARC formats and fields, to an FRBR model. Their research shows a ratio of 1:1.5 between expression and manifestation in the works included in the study.

Bennet, Lavoie and O'Neil [4] applied an FRBR model to WorldCat, assuming that each bibliographic record there described a manifestation. They found that the average work in WorldCat has about 1.5 manifestations, indicating that for the most part, works in WorldCat are small, single-manifestation entities. In fact, 78% of all works there consist of a single manifestation, and 99% have seven manifestations or less, while only about 1%, about 320K works, have more than seven manifestations. This result, however,

in no way should understate the potential of FRBR. Consider, for example, applying FRBR to an average Borders bookstore which contains 150k books or items in FRBR terms. In turn, those items can be traced back to a proportionately smaller number of manifestations, and, continuing up the hierarchy of entities, to an even smaller number of expressions and many fewer works. Therefore, the number of works represented in Borders will be some small fraction of 150k. This clearly suggests some benefit in applying FRBR to a small segment of the library catalog, i.e. the largest works.

Berg [5] compared FRBR with Taniguchi's expression-prioritized model and pointed out some problems associated with FRBR, including: (1) the likely confusion of attributes of manifestations with those of expressions; (2) an expression resulting from a combination of two existing works may need the creation of an extra new work to have a mapping relationship with said expression, as shown in Figure 1(W1 W2 W3 are works, E1 is an expression); (3) the need to define work and expression more precisely; and (4) the number of revisions an expression would go through before it is entitled to be referred to as a new expression.

Bennett, Lavoie, and O'Neil [4] defined three classes of works: (1) an elementary work is a work with a single expression and a single manifestation; (2)

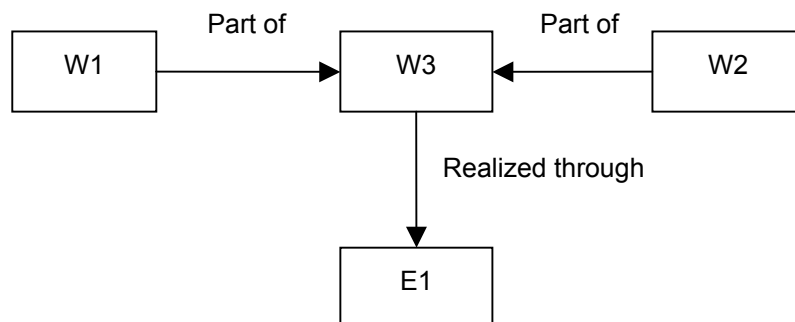


Figure. 1 Relationships between Work and Expression. [5]

a simple work is a work with a single expression but multiple manifestations; (3) a complex work is a work with multiple expressions of its intellectual or artistic content. Accordingly, because of multiple manifestations, theses belong to the second category. Berg [5], in the study mentioned above, classifies theses as simple works. In other studies, theses are treated as simple works since they basically exhibit one-to-one work-expression relationships. The following section details the application of FRBR to a thesis collection at the National Central Library of Taiwan.

## FRBR IMPLEMENTATION ON THESES COLLECTION AT NCL

FRBR's entities have been divided into three groups. The group 1 comprises the products of intellectual or artistic endeavour that are named or described in bibliographic records: *work*, *expression*, *manifestation*, and *item*. The group 2 comprises those entities responsible for the intellectual or artistic content, the physical production and dissemination, or the custodianship of such products: *person* and *corporate body*. The group 3 comprises an additional set of

entities that serve as the subjects of intellectual or artistic endeavour: *concept*, *object*, *event*, and *place*. This study first builds a mapping method for thesis metadata to FRBR three group entities. And then, we define relationship between three group entities. Finally, to establish FRBR bibliographic database for thesis..

## An FRBR Analysis of Theses Entities and Attributes

### Analysis of Theses Work Entity

Theses most often are works with singular expressions; therefore, a one-to-one relationship typically exists for work-expressions. The media manifestations of theses collected at NCL include paper hardcopies, image-files, internet web forms, and e-files. Electronic files are the original files written by research students, and NCL transformed those e-files into PDF, watermarked and coded them, so that they can be downloaded. Also, theses are kept in image-files for preservation. Paper hardcopies and internet web forms are kept at NCL and alma mater universities. Figure 2 below shows the entity-hierarchy of those theses:

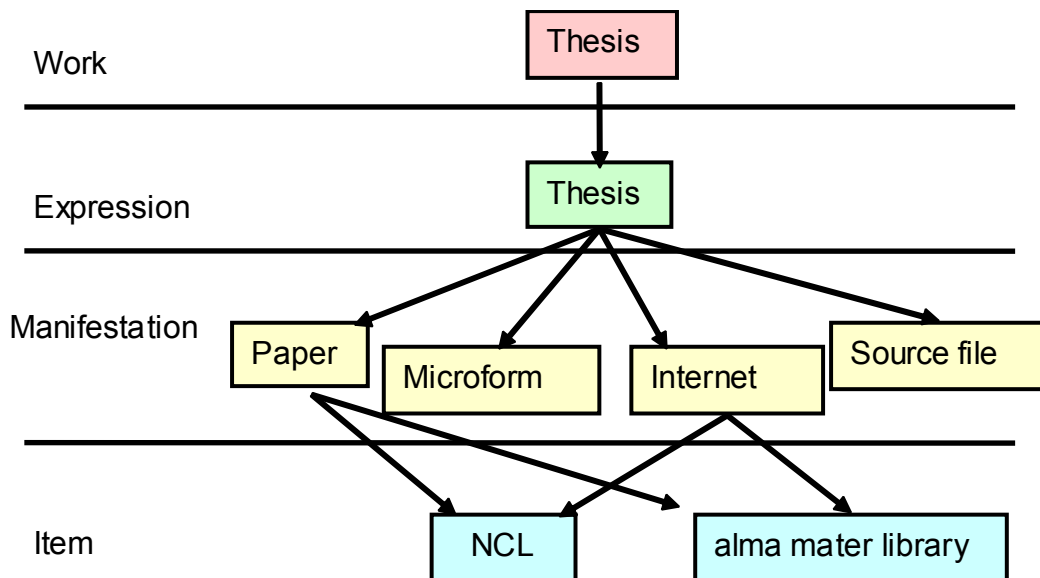


Figure. 2 Entity-hierarchy of theses

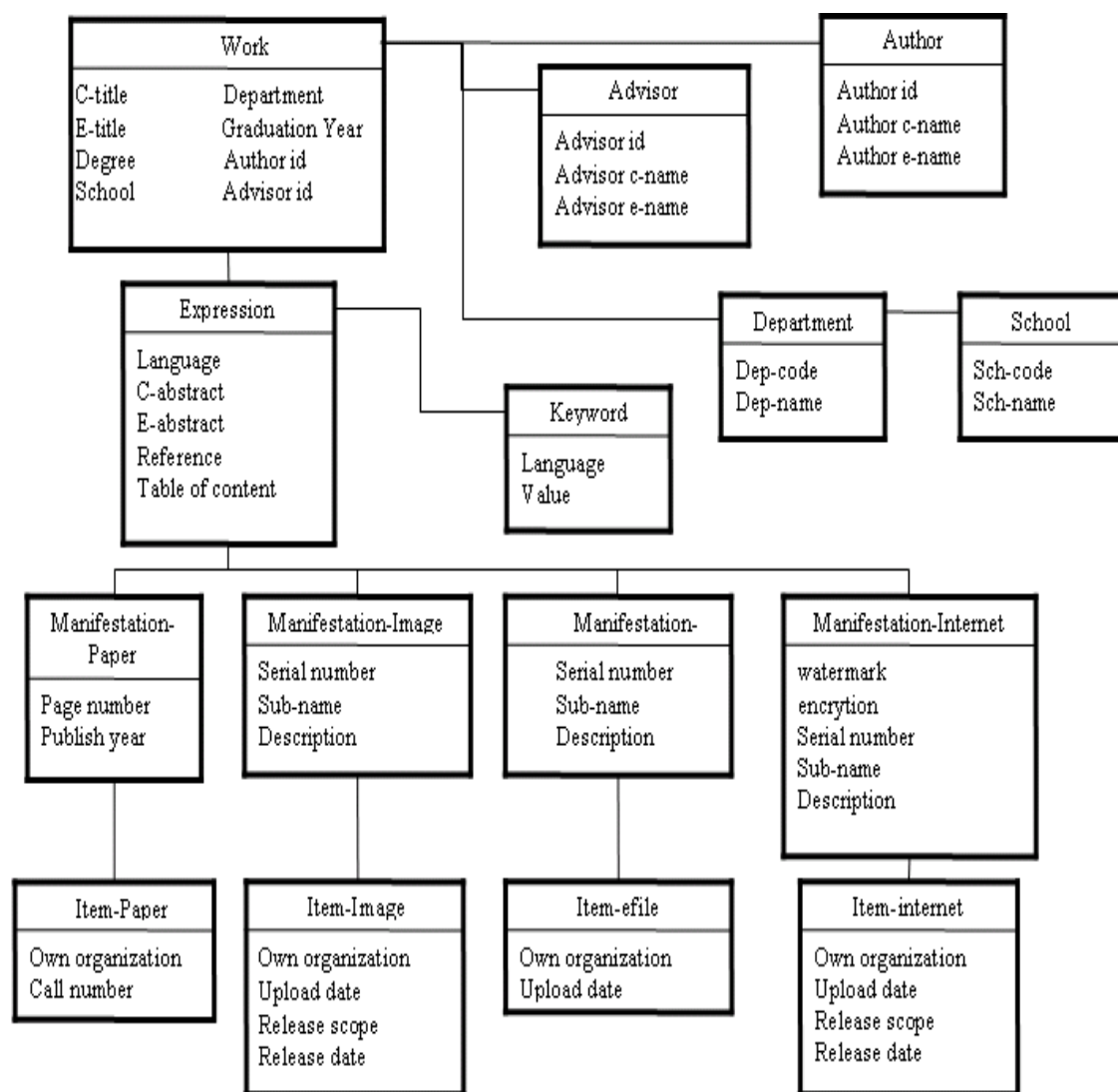


Figure. 3 Entity-attributes for thesis

## Mapping of Theses Interpretive Information to FRBR Attributes

After group 1 entity relationships are devised, we mapped theses interpretive information to FRBR attributes, to establish groups 2 and 3 entities and relationships. Group 2 entity-relationships refer to work in relationship to person and corporate body. Person entities include students and their advisors; corporate body entities include institutions, departments and NCL. Group 3 entities are those keywords entered by student-authors themselves. Figure 3 shows theses entity attributes.

### FRBR Defined Relationships

The fundamental distinctiveness of FRBR is the relationships between entities. Through relationships specified, users can access information with the least inconvenience. FRBR defines the following categories of entity-relationship (E-R).

1. Group 1 entities and primary relationships. The entities in the first group are work, expression, manifestation, and item. Their relationships are defined: work is realized through expression; expression is embodied in manifestation; manifestation is exemplified by item.
2. Group 2 entities and responsibility relationships. The entities in the second group are defined as person and corporate body. The relationships here are defined as follows: created-by, the relationship between work and group 2 entities; realized-by, the relationship between expression and group 2 entities; produced-by, the relationship between manifestation and group 2 entities; and owned-by, the relationship between item and Group 2 entities.
3. Other relationships among group 1 entities. For example, two works may relate to each other by their logical succession; that is, one is successor to the other.

4. Work relating to any entity in any group, including group 3 with its entities of concept, object, event, and place. Any entity and work can relate to each other by the relationship 'have-a-subject'.

The first two categories as described above are relationships inherent in FRBR itself as applied, while the last two categories are logical relationships to be possibly established when entities are identified.

For an FRBR application to theses, the first category of E-R is rather simple, while the second, third and fourth E-R would yield prolific relationships. In analyzing theses works, the following relationships appear.

1. Relationships in Group 1 entities: relationships between works (Figure 2).
2. Relationships between Group 1 entities and Group 2 entities: relationships between works and persons and corporate bodies (Table 1). Such relationships connect research students and their academic supervisors. To facilitate lookups, the study includes academic supervisors as an entity in Group 2. Accordingly, a new addition to the relationships is thesis supervision. Moreover, student-professor relationships are created.

Table 1. *Relationships between Group 1 entities and Group 2 entities for Thesis*

Entity(Group 1)	Relation	Entity (Group 2)
Work	is created by	Author, Advisor
Expression	is realized by	Author, Advisor
Manifestation	is produced by	Author
Item	is owned by	Own Organization

3. Relationships between Group 1 entities and Group 3 entities: relationships between works and concepts, as created by relating keywords used in works with Group 3 entities.

4. Relationships arising from references cited in works: reference is an important attribute of thesis works, and research literature findings would be greatly increased by using the reference relationship. Therefore, reference relationship is added.
5. Successor relationships: These usually contain suggestions for follow-up studies to come, and thesis works can be quite meaningfully associated by successor relationships.
6. Complement relationships: This relationship occurs when several theses have been conducted to study one major concept from various perspectives dealing with different aspects of the central idea. This is not uncommon when several students have produced theses under the supervision of the same academic advisor, with each one concentrating on a different area in their field of study. Accordingly, thesis works arising from similar academic interests and parameters would reflect such complement relationships.

This study is the first of a two phase research scheme, focusing on the first 4 (1-4) relationships. The second phase will examine the last 2 (5-6) relationships, involving more complicated logical relationships that will require in-depth knowledge in ontology and data mining. The results from our first phase study are presented below.

## **FRBR Experiment on Theses at NCL**

### **Theses Works Used**

The study's universe included theses at NCL with the word "Library" in their titles. A total of 572 thesis works were used, and the academic institutions involved included the Departments of Library and Information Science in the following universities in Taiwan: National Chung-Hsing University, National

Taiwan University, National Chengchi University, TangKang University, Fujen University, and National Normal University. This represents a well-balanced approach to data selection and has in the process included other theses available via a national indexing network.

## **Establishing FRBR Bibliographic Database**

Two steps are involved in establishing FRBR E-R Database. First, bibliographic records from NCL are standardized and normalized so that they are compatible with the target tables. Second, mapping to FRBR is performed by algorithmic procedures designed specifically for the purpose. The following tables were established for this study:

1. Work-Expression Table: Work and Expression have a one-to-one relationship; therefore they are combined into one table.
2. Supervision Table: This lists the many-to-many relationships between academic supervisors and works produced by their research students.
3. Keyword Table: This lists the many-to-many relationships between works and keywords they use.
4. Reference Table: This lists relationships between reference works and referring works.
5. Student-Professor Table: This lists relationships between student and their academic supervisors.

## **Results from the Study**

By using the algorithm designed for this study, the following three results have been produced, displaying their respective E-R.

1. Result Displaying Work Relationships and Selections. Figure 4 shows a certain thesis selected, its works, expressions, manifestations, and items. Users can pick the particular expression, manifestation and items of their choice.

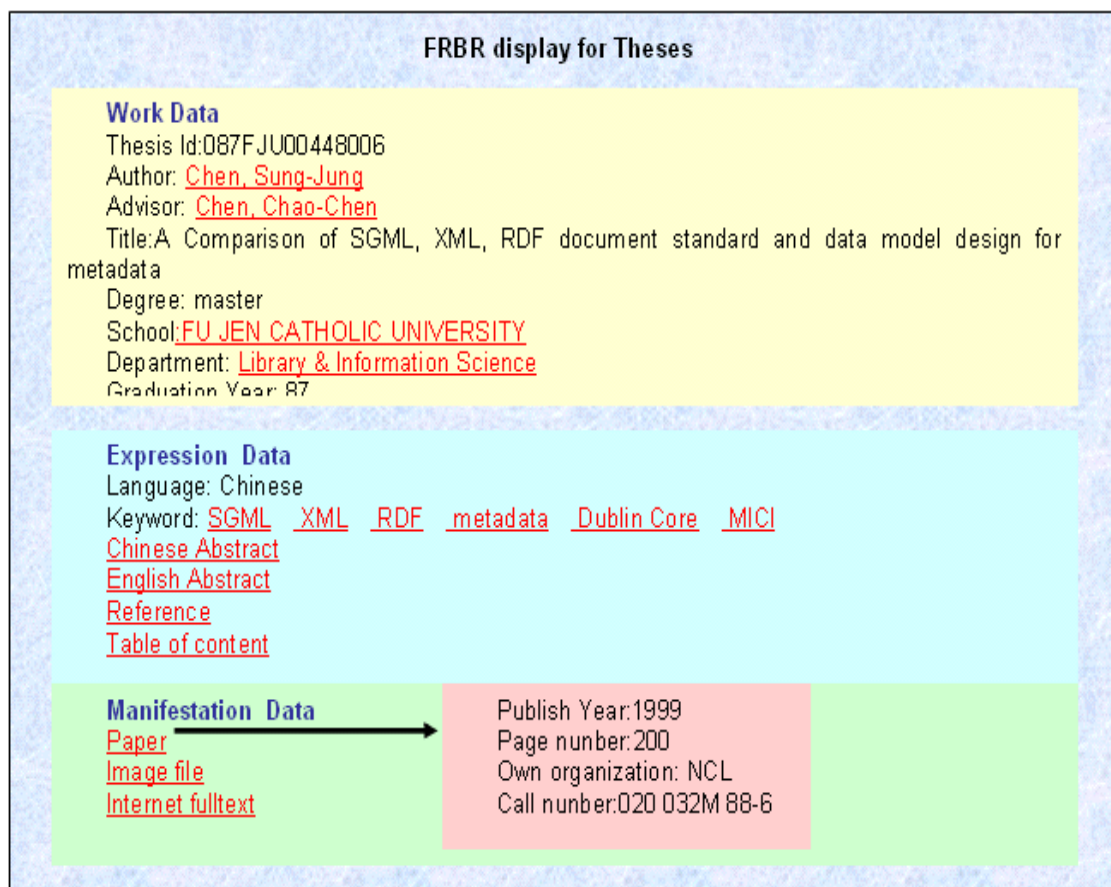
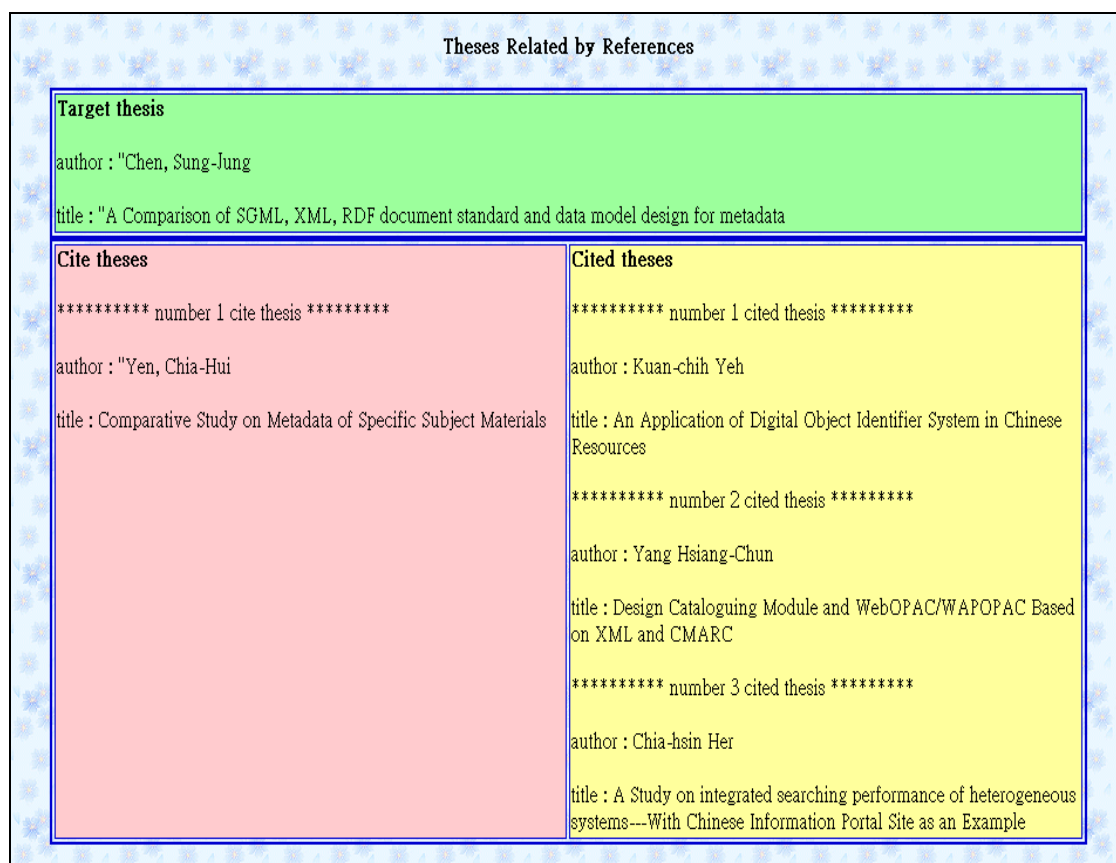


Figure. 4 FRBR Display for theses

2. Result Displaying Theses Related by References Used. This result would help users to find related theses by references cited in them. Figure 5 shows in the upper half a specific thesis

work, while its lower left lists the theses this work cited and its lower right lists theses that cited it. Through this E-R, users can find the predecessors and successors related to the work of interest.





*Figure. 5* Theses Related by References

In the theses works accounted for, there are 258 theses which have reference citations, all of them being works produced after 1996. The total number of theses works produced in Taiwan was 1551, including 573 relating to library science, and 978 relating to other related academic fields, the ratio between them being roughly 3:5. Further analysis reveals that 978 references cited from related fields have the following statistics: 262 from education, 149 from business administration, 136 from information technology, 57 from media and journalism, and 34 from law. The above data shows the

relatedness of library science to these subjects in descending order.

### 3. Result Displaying Professor-Student Relationships.

This FRBR experimental study relates professors and students as persons to persons. Looking up a person would result in persons relating to her/him as student or professor. Moreover, the algorithm used would trace backward or forward until all the generations and persons existing in such professor-student relationships are shown. As Figure 6 shows, the first generation is root node and second generation is sun node, and so on.

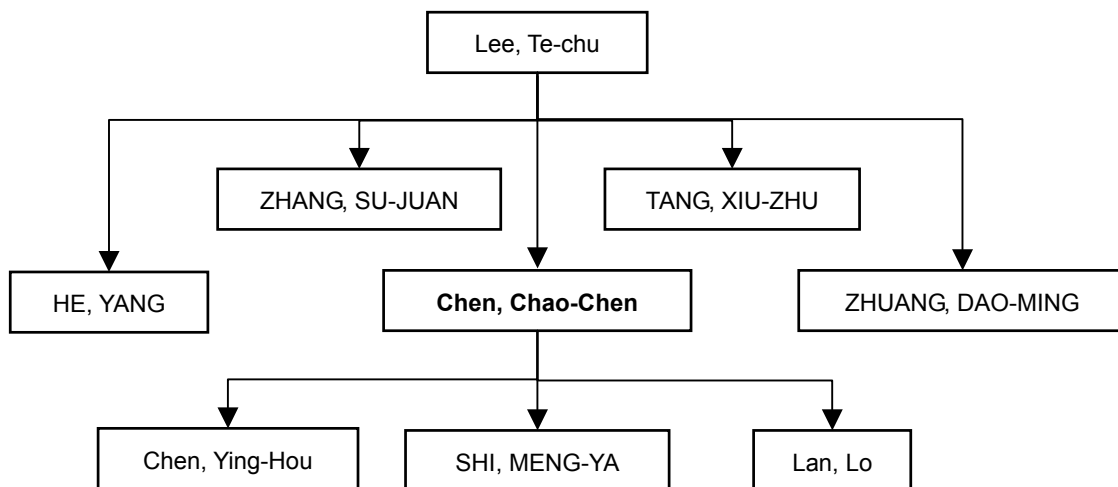


Figure. 6 Professor-Student Relations

## CONCLUSIONS AND DISCUSSION

Based upon the results presented, this study has reached the following conclusions.

1. Thesis works differ from other works in that they have unique attributes unaccounted for by FRBR. Due to their relative independence, thesis works do not share the Group 1 entities FRBR defines and lack most of the relationships contained therewith. Furthermore, E-R as defined by FRBR needs much additional human judgment, rather than automatically obtaining information from existing bibliographic records.
2. Lacking in the FRBR model, citations and references E-R are vital parts of theses and other research oriented works. This study establishes the crucial nature of E-R for thesis citations and references. From such E-R, users can link up with theses related to the targeted one. Theses in library science appear to cite mostly from theses in education, an interesting bit of information only available with E-R for citations and references.
3. In addition to citation E-R, this study establishes two other E-R: student-professor and predecessor-successor E-R for theses works, enriching FRBR to function more effectively for thesis works.
4. It is evident that theses works do not have the complete E-R set FRBR has for Group 1. Rather, they seem to yield more E-R related to Group 2 and 3 than FRBR proposes. Therefore, in the future, FRBR may need to include a development of more Group 2 and 3 E-R, so that it may serve the individual needs arising from specific contexts and environs.
5. The bibliographic records used by this study were not originally coded in MARC format. Rather, their format is interpretative and specifically designed for thesis works. The mapping between them and FRBR therefore mostly involves work and expression. This is in line with the findings of Chen, Y.N., Lin,

S.C., & Chen, S. J., [7] in their interpretative format done for library materials. However, Bennet, Lavoie, and O'Neil [4] used MARC bibliographic records for their study, resulting in greater occurrence of manifestation. A comparison between interpretative and MARC formats reveals that the former would involve work and expressions while the latter manifestations.

6. This study devised four levels/categories of E-R in user interface, so that the user can have a better knowledge structure to work than traditional cataloguing and indexing methods. This seems to be rather helpful when dealing with a large database, but when dealing with a relatively small database, it would not be worth the effort to build additional E-R levels.
7. The work is alone in the traditional cataloguing and indexing methods, and then the FRBR mode construct various knowledge network for work. This study use reference relationship and professor-student relationship to build work network. Beside, the work of FRBR provides user with many kinds of media for theses and let user easy to find the content of theses.

To conclude our discussion, the study has contributed to showing certain FRBR strengths and conveniences. It has put into practice FRBR knowledge structure in helping users to achieve maximum effectiveness and efficiency in navigation through bibliographic universes. Thus, FRBR potentiality is shown to create positive impact upon library users in this age characterized by an information explosion. In addition, the study finds that FRBR mode is not enough the relations for theses. And then, the study enhances two relations for the FRBR mode of the theses

## NOTE

[1] Chang, H. C. & Lin, S. N. (2004). The Development of the Functional Requirements for Bibliographic Records. *Bulletin of the Library Association of China*, 73, 45-71.

[2] Hickey, T. B., O'Neill, E. T. & Toves, J. (2002). Experiments with the IFLA Functional Requirements for Bibliographic Records (FRBR). *D-Lib Magazine*, 8(9). Retrieved April 26, 2006, from the World Wide Web: <http://www.dlib.org/dlib/september02/hickey/09hickey.html>

[3] Hegna, K. & Murtomaa, E. (2002). *Data mining MARC to find: FRBR?* Retrieved April 26, 2006, from the World Wide Web: <http://folk.uio.no/knuthe/dok/frbr/datamining.pdf>

[4] Bennett, R., Lavoie, B. & O'Neill, E. (2003). The concept of a Work in WorldCat: An application of FRBR. *Library Collections, Acquisitions, and Technical Services*, 27(1). Retrieved October 20, 2005, from the World Wide Web: [http://www.oclc.org/research/publications/archive/2003/lavoie\\_frbr.pdf](http://www.oclc.org/research/publications/archive/2003/lavoie_frbr.pdf)

[5] Berg, E. S. (2004). *Implementing FRBR: A Comparison of two relational models: IFLA's FRBR model and Taniguchi's expression-prioritized model*. Unpublished master's thesis, Oslo University College, Norway.

[6] Day, M. (1998). *Data models for metadata: Some issues for the Dublin Core initiative (draft)*. Retrieved March 1, 2006, from the World Wide Web: <http://www.ukoln.ac.uk/metadata/data-models/draft-report.html>

[7] Chen, Y. N., Lin, S. C. & Chen, S. J. (2002). An application practice of the IFLA FRBR model: A metadata case study for the National Palace Museum in Taipei, Proceedings of the 65<sup>th</sup>. *Annual Meeting of the American Society for Information Science*, 39, 181-193.