Applying UML 2.0 to Design a Botanical Document Warehouse

Rong-Jyue Fang\textsuperscript{1}, Howard Lo\textsuperscript{2}, Chien-chung Lin\textsuperscript{3}, Yu-Chen Weng\textsuperscript{4}
\textsuperscript{1}Chair Professor, Department of Information Management, Southern Taiwan University of Technology, Taiwan
\textsuperscript{2,4}Graduate student, Department of Industrial Technology Education, National Kaohsiung Normal University, Taiwan
\textsuperscript{3}Professor, Department of Business Administration / Graduate Institute of Management, Meiho Institute of Technology, Taiwan

\textbf{ABSTRACT:} The Unified Modeling Language (UML) is the de facto standard for modeling object-oriented software and system analysis. The four plus one View Model describes the information system framework using five concurrent views, each of which depicts a specific view of the system. UML consists of various graphical notations, which capture the software designer’s architecture decision. The nine diagrams can assist system designer and analyzer to sketch the blueprint of information system easily.

Data warehouse is the latest must-have marketing weapon in Business domain. It provides architectures and tools for users to systematically organize, understand, and use their data to make strategic decision. It is also useful for information retrieval in pedagogical settings.

In this study, we constructed the framework of the prototype of multi-dimensions botanical documents inquiry system (MDBDIS) with the notation and methodology of UML. The empirical results of this study has showed the high recall and precision rates, the recall rate of LeaveMargin = “Serrate” and florescence=“pink” is 0.83, and the precision rate of LeaveMargin = “palmatelylobed” and Phyllotaxy =“opposite” is 075.

In this paper, the finding which has provided evidence of an empirically applied UML methodology to sketch the blueprint of data warehouse is a useful new avenue for system designer and analyzer.

\textbf{Keyword:} UML, Data Warehouse, Document Warehouse, Knowledge Management, Botany

\section{Introduction}

Since its introduction by Grady Booch, Jim Rumbaugh, and Ivar Jacobson in 1997, the Unified Modeling Language (UML) has capture the interest of software designers and system analyzers\cite{2}. The UML has became the de facto standard for modeling object-oriented software and system analysis\cite{3,4}. Fowler distinguishes three types of UML usage, such as sketch, blueprint, and programming language\cite{5}. The four plus one View Model describes the information system framework using five concurrent views, each of which depicts a specific view of the system\cite{1}. Grady Booch, Jim Rumbaugh, and Ivar Jacobson, the so-called Three Amigos of software engineering redefined the 4+1 View Model and exploited several notation and diagram to describe the framework of information system. UML consists of various graphical notations, which capture the static system architecture, system component behavior and system component interactions\cite{6}. Use Case diagram, Case diagram, and Component diagram captures the static system architecture, Statechart diagram describes the system component behavior, Collaboration and Sequence diagram depicts the system component interactions. In many cases, UML have been exploited to systematically represent the frameworks of an information system. Its importance has increased with the 4+1 view model-driven architecture methodology\cite{2}.

Data warehouse are widely applied in business domain to assist the CEO’s (Chief Executive Officer) decision making. Data warehousing provides architectures and effective tools for CEO to systematically organize, understand, and integrate the data from various kind of databases to support their decision making strategy\cite{7}. The data warehouse design process is vital and should be supported by an systematic method. There is a lack of a general data warehouse design method to specify the various design steps\cite{8}. Considering the need for a systematic data warehouse method, this paper presents a UML-based methodology for the design of data warehouse.

Basically, the content in the data warehouse are structured data extracted from operational systems, and can be used for data mining and knowledge
discovery. Besides structured data, a large percent of information are stored in a static unstructured format in the data collection. According to the study of Sullivan, there were twenty percent of information stored in the format of structural data, such as relational database; eighty percent of information were came from unstructured data, such as text, and document[9]. In this study, we applied the Unified Modeling Language to design a document-based data warehouse to fulfill the users’ information need. Since botanical learning activities are ubiquitous in the subject of Science and Technology of the K12 Curriculum[10]. In this study, we applied UML notation to capture the data warehouse design decision, and a multi-dimensions botanical documents inquiry system (MDBDIS) prototype has been developed.

In the following section we will introduce the four plus one view of UML, and notation of nine diagrams. The third section will introduce the architecture of data warehouse, and the prototype of MDBDIS. The fourth section will discuss the efficiency and effectiveness of MDBDIS. The last section will be the conclusion of this paper.

2 The Unified Modeling Language

Object-oriented(OO) technology has profoundly changed the landscape of software design and development[11]. According to Jonson’s survey, seasoned software developer are convinced that Object-Oriented systems development (OOSD) delivers distinct advantages over conventional structured or process oriented approaches, such as easier modeling, increased code reuse, higher system quality, and easier maintenance[12, 13]. In 1997, Booch, Rumbaugh, and Jacobson, the three amigo of software engineering integrated their object-oriented methods , and redefined the Kruchte’s 4+1 view model into the concept of UML 1.0. The advent of UML has sparked the continued growth and acceptance of OO technology[13].

2.1 The Four Plus One Views

During the early 1990s, there were various concepts related to the object-oriented methods. Prior to 1994, there were many competing visual modeling languages and methodologies on the market. In 1995, Kruchten first introduced the 4+1 view model in his paper. This model used five concurrent views to organize an information system architecture. The Fig. 1 depicted the logical view, process view, physical view, and development view. The descriptions are as follow.

- The logical view: It describes the object model view of the system design.
- The process view: Describes the concurrency and synchronization aspects of system design.
- The physical view: Describes the mapping of the software onto the hardware and reflects its distributed aspect.
- The development view: Describes the software’s static organization in its development environment.

In 1997, Booch, Rumbaugh, and Jacobson proposed the UML 1.0, and accepted by Object Management Group (OMG). UML is a visual modeling language, consisted of notation and textual components to describe the Object-Oriented systems development. The three amigo redefined Kruchten’s 4+1 view model into Fig. 2. The new 4+1 view model is a blueprint of an information system. The system analyzer can apply this model to extract the design information, and sketch the system blueprint. The description of 4+1 view of UML are as follow:

- Use Case View: Describes the system function from the view of users. Use case diagrams of UML are used to view a system from user’s perspective via a set of discrete activities or transactions.
- Design View: Describes the static system architecture, system component behavior. Class diagrams and object diagrams form the design view of the system.
- Process View: Describes dynamic behavior of a system, such as thread and process. The different diagrams such as the state diagram, activity diagram, sequence diagram, and collaboration diagram are used in this view.
- Implementation View: Describes the design view of an object with modules or component.
- Deployment View: Describe the deployment scenario.

Fig. 1 Kruchten’s 4+1 View Model

Fig. 2 The 4+1 View Model of UML
modules in the system. The deployment diagram can be used to identify the deployment modules for a given system.

2.2 The UML Diagram

The unified modeling language provides nine different diagrams to assist systems analyzer to sketch the framework of information system. In this paper, we exploited the use case diagram, activity diagram, and class diagram to construct the framework of multi-dimensions botanical documents inquiry system. The nine diagram are described below:

- Use case diagram: A use case diagram describes a sequence of actions from the user’s perspective. Use case is drawn as a horizontal ellipse represent the botanical document warehouse, and actor represent the users, such as naïve user, students, parents, and teachers. Fig. 3 shows the use case diagram.

- Activity diagram: An activity diagram is typically used for business process modeling. With this diagram, system analyzer can model a single use case or usage scenario from logic view. Fig. 4 shows the activity diagram.

- Class diagram: A class diagram shows the classes of the system, their interrelationships such as inheritance, aggregation, and association, and the operations and attributes of the classes. Fig. 5 shows the class diagram.

- Object diagram: An object diagram is useful for exploring the real information system world. It can model the related object within the system.

- Sequence diagram: A sequence diagram models the flow of logic within your system in a visual manner, and is commonly used for both analysis and design purposes.

- Collaboration diagram: A collaboration diagram depicts the same information as sequence diagram via different method. System analyzer can understand the work flow among the objects through this diagram.

- State chart diagram: UML 2.0 rename this diagram as state machine diagrams. This diagram describes the various states that an object may be in and the transitions between those states.

- Component diagram: A component diagram model the information system architecture, describes the inter-relationship among the components, and implementing interface.

- Deployment diagram: A deployment diagram depicts the system networks and the layout of components via a static view[14].

With UML diagram and notation, system developer can easily capture the system framework to sketch the blueprint of their software system, even implement the blueprint into real work information system.

Fig. 3 Use case diagram

Fig. 5 Class diagram

Fig. 3 Activity diagram
3 Constructing an Multi-dimensions Inquiry System document system

In this study, we apply the 4+1 view model and UML diagram to sketch the framework of our empirical data warehouse, and integrate the concept of data warehouse and document warehouse into our Multi-Dimensions Botanical Documents Inquiry System (MDBDIS).

3.1 The concept of Data Warehouse

Data warehouse is the latest must-have marketing weapon in Business domain. It provides architectures and tools for users to systematically organize, understand, and use their data to make strategic decision. It is also useful for information retrieval in pedagogical settings.

3.1.1 What is a Data Warehouse?

The construction of data warehouse, which involves data cleaning and data integration, can be viewed as an important preprocessing step for data mining[7]. Data warehousing provides architectures and tools for decision-makers to systematically organize, understand, and use their data to make strategic decisions. Data warehouses have been defined in many ways, making it difficult to formulate a rigorous definition. According to W.H Inmon, a lead scholar in data warehouse domain, “A data warehouse is a subject-orient, integrated, time-variant, and nonvolatile collection of data in support of managements’ decision making process”[7, 15].

3.1.2 The concept of Document Warehouse

According to the report of Survey.com (http://www.survey.com), 20% of information was stored in the form of structural data, like relational database; 80% of information was came from non-structural data, such as text. As we delve into the realm of text-oriented business intelligence, we will quickly see two parallels with more traditional numeric business intelligence operations. First, in both cases we are dealing with large quantities of information, since 80 percent of all business information is textual, documents warehouses will have to deal with massive amounts of raw data[9]. Since there are bountiful botanical text resources on the Internet, we need a good tool to search and retrieve the related information from the Internet. According to Sullivan[9], “If the web is the mine from which we extract raw material, document warehouses are the mills and refineries through which we turn text resources into usable business intelligence and distribute it to the ultimate customer.” Applying UML methodology to construct the framework, and integrating the concept of document warehouse into the multi-dimensional botanical inquiry system is the aim of this study.

3.1.3 The framework of MDBDIS

In this study, we use the UML diagrams and notations to capture the users’ information need and extract the expertise from several botany experts. The UML use case diagram shows in Fig. 6. Based on these concepts, we applied the data warehouse concept to construct a multi-dimensions botanical documents inquiry system to assist the naïve users. Users can observe the features of a plant such as Phyllotaxy, Leaf shape, Apex, Base, Margin, Inflorescence as inputs of our Multi-Dimensions Botanical Documents system. With the mobile device like 3G mobile phone or PDA, users can easily found the information related to the specific plant anytime, anywhere.

![Use Case Diagram](image)

Fig. 6 Example of use case diagram related to the specific plant anytime, anywhere.

3.1.4 The prototype of MDBDIS

We developed the MDBDIS system on an IBM xSeries 345 Server with 2 Intel Xeon processors, 4 SCSI 72GB hard disk driver, and 2GB of DDR memory. The platform was Linux series such as RedHat 9.0. We implement our system with Java language, JavaServer Pages, and MySQL database. Java language is a cross-platform compatibility and write-once-run-anywhere computer language, it is also free on the Internet. MySQL is the same as free as Java. With these two powerful tools, we can develop our MDBDIS easily, and don’t worry about the budgets.

4 Efficiency and effectiveness of MDBDIS

When considering query performance evaluation, we should first consider the inquiry task that is to be evaluated. Efficiency and effectiveness is the most important part of evaluation work. We adopted the recall rate and precision rate to assess MDBDIS system.
4.1 Recall rate
Recall rate is the fraction of the relevant botanical documents which has been retrieved. The Equation (3) depicts the calculation of recall rate. In our MDBDIS system, the recall rate of LeaveMargin = "Serrate" and florescence="pink" would be 0.83, for 5 relevant documents had been retrieved, and there 6 relevant documents in document warehouse.

4.2 Precision rate
Precision rate is the fraction of the retrieved botanical documents which is relevant. The Equation 4 will indicate the precision rate. The precision rate of LeaveMargin = "palmately lobed" and Phyllotaxy = “opposite” is 0.75. There are 3 relevant documents had been retrieved and 4 documents had been retrieved.

5 Conclusion
In this paper, We applied the UML methodology to sketch the blueprint of multi-dimensional botanical document inquiry system. The use case diagram is very useful for system analyzer to capture the users’ information need and the expertise from the expert of botany domain. The activity diagram can assist the system developer to implement the real system. The class diagram help designer to describe the attribute of every class.

In this paper, we integrated the concept derived from data warehouse and document warehouse into the prototype of MDBDIS. With the multidimensional advantage, The MDBDIS can provide an effective tool for naive user to fulfill their information need.

The empirical result of this study has showed the high recall and precision rates, the recall rate of LeaveMargin = "Serrate" and florescence="pink" is 0.83, and the precision rate of LeaveMargin = “palmately lobed” and Phyllotaxy = “opposite” is 0.75. With MDBDIS, the naive users can easily key in the features of a plant, such as Phyllotaxy, Apex, Base, and LeafMargin, as a input of the query, then the system will narrow down the number of relevant documents to meet user’s information need.

In this paper, our finding have provided evidence of an empirically applied UML methodology to sketch the blueprint of data warehouse is a useful new avenue for system designer and analyzer. In future research, We want to extend our research to the other specific domain such as entomology, zoology.

References: