Republic of Iraq
Ministry of Higher Education
and Scientific Research
Al-Nahrain University
College of Science

Database Personalization in E-Learning

A Thesis
Submitted to the College of Science, Al-Nahrain University
In Partial Fulfillment of the Requirements for
The Degree of Master of Science in Computer Science

By
Shelan Sabah Ahmed

(B.Sc. 2005)

Supervisor
Dr. Ban N. Al-Kallak

2008 ***1429***

بِسْمِ اللهِ الرَّحْمَنِ الرَّحِيمِ

يُؤتِى الحِكمَةَ مَن يَشَآءُ ۚ وَمَن يُؤتَ
الحِكمَةَ فَقَدأُوتِىَ خَيرا ًكَثِيراً وَمَا يَذَّكَّرُ
إلا أُوْلُوا الأَلبَابِ

صَدَقَ اللهُ العَليُّ العَظيم

البقرة ٢٦٩

Acknowledgement

Before every thing and after thanking of ALLAH, I would like to express my sincere gratitude and appreciation to my supervisor **Dr. Ban N. Al-Kallak** for her valuable guidance, supervision and untiring efforts during the course of this work.

I would like also to express my thanks to the college of science, dean of the collage, **Dr. Taha S. Bashaga** the head of computer science department and all staff and friends for the continuous support and encouragement during the period of my studies.

My special thanks to my husband, for his continual love, support, and patience.

Finally, my special thanks to my family for their continuous support and encouragement during the period of my studies.

Shelan

Dedication

I dedicate my work to...

My Parents

My Husband

My Brothers

My Sisters

And

My Friends

With love and appreciation for their support and encouragement.

Shelan

Abstract

Personalized support for learners became even more important when e-learning take place in distributed database system.

This research provides access to learning resources in a distributed e-learning database and it intends to adapt the system to the learner’s needs (personal needs) so that each learner gets the idea that the system was created just for him and knows what he likes. The personal needs are determined in different ways like the current request (search), the previous resource requests (personal search about the previous requests) or the education level of the learner (syllabus). The personal search methods are implemented by creating explicit and implicit user profiles to store the learner's information and his previous resource requests (interest). The personalization techniques that are used in this research are content-based filter and rule-based filter which have rely heavily on user profiles.

The proposed system ensuring security in terms of authentication (identify the user and his permission), integrity (the data is not getting changed or corrupted), and confidentiality (other people should not be able to see other user’s personal information in the user profile).

The proposed system architecture is Two-Tier Multiple-Client / Multiple-Server model which consist of two data servers and multiple clients. Each server used to store a fragment of the data. The servers are located in locations where there data are most widely used to decrease the size of the data stored in each server, decrease the load over the distributed database management system and most operations are performed locally.

Table of Contents

Chapter Four: Design and Implementation of the Proposed System

Chapter Five: Conclusions and Suggestions for Future Works

List of Abbreviations

CBL	Computer Based Learning
CBT	Computer Based Training
CD-ROM	Compact Disk-Read Only Memory
CS	Communication System
DBMS	Data Base Management System
DDB	Distributed Data Base
DDBMS	Distributed Data Base Management System
DDBS	Distributed Data Base System
DDL	Data Definition Language
DML	Data Manipulation Language
Dom	Domain
E-Learning	Electronic Learning
E/R	Entity/Relationship
ID	Identifier
MBL	Multimedia Based Learning
NF	Normal Form
ODBC	Open Data Base Connectivity
PDEL	Personalization in Distributed E-Learning
SQL	Structured Query Language
TV	Television
WBL	Web Based Learning

جمهورية العراق
وزارة التعليم العالي و البحث العلمي
جامعة النهرين
كلية العلوم

اضفاء طابع الشخصية على قاعدة بيانات تعليم الكتروني

رسالة مقدمة الى كلية العلوم، جامعة النهرين و هي جزء من متطلبات نيل
شهادة الماجستير في علوم الحاسوب

من قبل

شيلان صباح احمد

بكالوريوس
٢٠٠٥

المشرف

د. بان نديم الكلاك

٢٠٠٨ ١٤٢٩

الخلاصة

أصبحت المساعدة الشخصية للمتعلمين مهمة لدرجة أكبر عندما اخذ التعليم الالكتروني مكانهُ في نظام قاعدة البيانات الموزعة.

يزودنا هذا البحث بالوصول الى مصادر التعليم في قاعدة بيانات التعليم الكتروني الموزعة ويحاول تكييف النظام طبقاً لاحتياجات المتعلم (احتياجات شخصية) لذلك كل متعلم تخطر له فكرة ان النظام مُعد من اجله فقط ويعرف مالذي يرغب فيه. إن الاحتياجات الشخصية تُحدد بطرق مختلفة مثل الطلب الحالي (بحث) او المصدر المطلوب سابقاً (بحث شخصي حول الطلبات السابقة) أو المستوى العلمي للمتعلم (منهجه الدراسي). ان طرق البحث الشخصي تُطبق بِخلق مَلف ظاهر و مَلف ضمني لخزن معلومات المتعلم والمصادرالتي طلبها سابقاً (المفضلة) لأن تقنيات إضفاء طابع الشخصية المستعملة في هذا البحث هما مرشِح أساسهُ المحتوى ومرشِح أساسهُ القاعدةَ التي تعتمد بشدة على محتويات الملف الظاهر والملف الضمني للمستخدم.

النظام المقترح يضمن اجراءات أمنية من ناحية التوثيق (يُعرّف المستعمل ورخصته) و النزاهة (البيانات لايمكن تغييرها او تحريفها) و الخصوصية (أشخاص آخرون لا يستطيعون رؤية المعلومات الشخصية لمستخدم اخر في ملفه الشخصي).

إنّ معمارية النظام المقترح هي نموذج مزدوج-الطبقات متعدد-التوابع / متعدد-الخدم الذي يتألف من خادمان للبيانات والعديد من التوابع. كل خادم يُستعمل لخزن جزء من البيانات و يُوضع في الموقع الذي يكون فيه اكثراستخداماً و ذلك لتقليل حجم البيانات المخزونة في كل خادم وتقليل الحمل على نظام إدارة قواعد البيانات الموزعة وأكثر العمليات ستكون موقعية تماماً مما يقلل المرور عبر الشبكة.

Chapter Five

Conclusions and Suggestions for Future Works

5.1 Conclusions

The conclusions that can be drawn from this work are listed as follows:

1. The system was used to provide personalized access to learning resources in a distributed e-learning database making the access to the resources faster and easer.
2. Different ways like the current request (search), the previous resource requests (depending on date or frequency) or the education level of the user (syllabus) were taken in to account to simplify personal search.
3. The analysis of the explicit and implicit user profile data is based on the static rule based filter and content based filter which provides the simple way to implement the personalization process.
4. The Two-Tier Multiple-Client/Multiple-Server model used in the proposed system has the benefits over the centralized system:
    - Allowing the partitioning of database relations, store each part in one server, and allowing a fragment of a given relation to be stored in to multiple servers to reduce the load over the DDBMS.
    - This model is economic because many small computers used as a server and its cost less than a single big one.
    - This system model is modular growth; the new modules (client or server) are easily added to the system.

5.2 Suggestions for future works

There are several ideas for developing the proposed PDEL system such as:

1. Support replication by making many copies for the relations and store each copy in one server, each group of clients connects to the nearest server to it and the updates at a primary server will automatically updates all the replicas of the database. These replicas may be used as backup's data storage.
2. Giving the manager a control to update the learner's account when the learner forgets it.
3. It also seems interesting to study a way to increase the system security.
4. Extend the system to be accessed from a web browser.

Chapter Four

Design and Implementation of the Proposed System

4.1 Introduction

Personalization in distributed e-learning (PDEL) is used to give the learners support in accessing the resources. It uses two methods to meet the learner's need. The first method is the *search* which returns the resources about the user's current request, and the second method is *personal search* which has three sub methods, the first one returns the resources about previous resource request depending on the latest date, the second one returns the resources about previous resource request depending on the frequency, these methods accomplished by using content-based filter, and the third method returns the resources about the user's syllabus which accomplished by using static rule-based filter. The personal search methods are implemented by creating explicit and implicit user profiles to store the learner's information and his previous resource requests (interest), because the personalization technique that is used in this research (content-based filter and rule-based filter) has rely heavily on user profiles.

The users of the e-learning may be *manager*, *authorized persons* or *learners*. The manager is a special type of the authorized persons which has full controls over the PDEL and has the permissions to create and update the authorize persons account and determine them controls by classifying them into groups. The authorized persons have controls (insert, delete, and update) over the resources and syllabus, and the learners have the permission to search about the resources only. The overall modules of the personalization in distributed e-learning are shown in Figure (4.1).

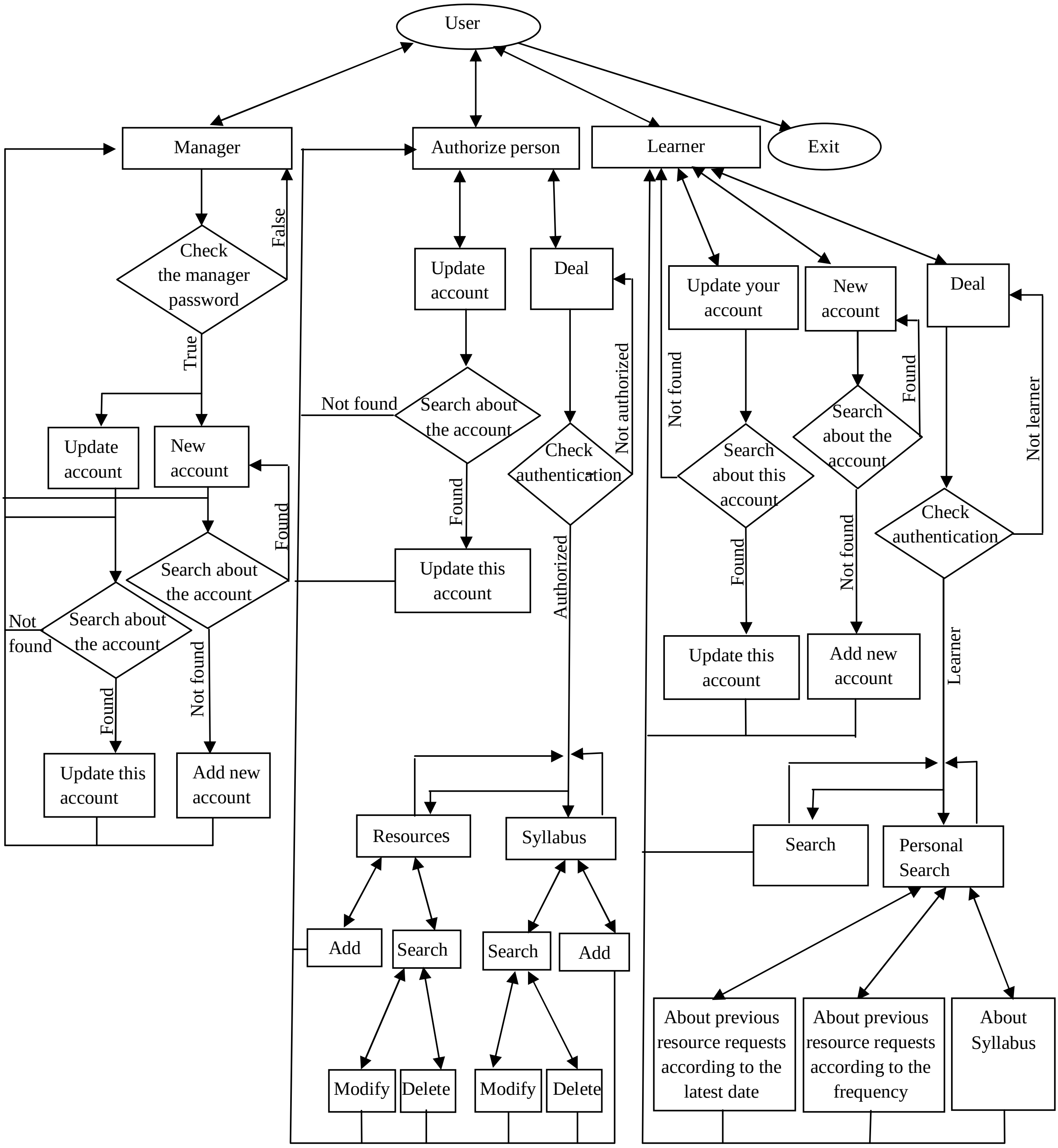


Figure (4.1) the overall modules of PDEL.

PDEL ensuring security in terms of authentication (identify the user and his/her permission), integrity (the data is not getting changed or corrupted), and confidentiality (other people should not be able to see one user's personal information in the user profile).

PDEL needs large data storage system so we use distributed database system. MySQL Server 5.0 database was chosen as a data base language, Microsoft visual basic 6.0 as programming language and MYSQL Connector/ODBC 3.51 as a connector between them.

4.2 The Proposed System Architecture

The architecture of the proposed system is Two-Tier Multiple-Client/Multiple-Server model. The relations of the PDEL are fragmented in to two parts, each part is stored at different site (location) to decrease the size of the data stored in each database and decrease the load over the DDBMS. These sites are called the servers and the PDEL application is called the client. The users have to deal with fragments but they need not to know about the location of these fragments because the system provides two transparencies: fragmentation transparency and location transparency. So that PDEL designed in Client/Server model which consists of nine computers two of them specified as database servers and the other defined as clients (Distributed System), each server may be defined as client and server. The databases, DDBMSs, and applications of the PDEL are of the same type so it is in homogenous distributed system. These computers connected via a switch as star network. Figure (4.2) shows PDEL architecture.

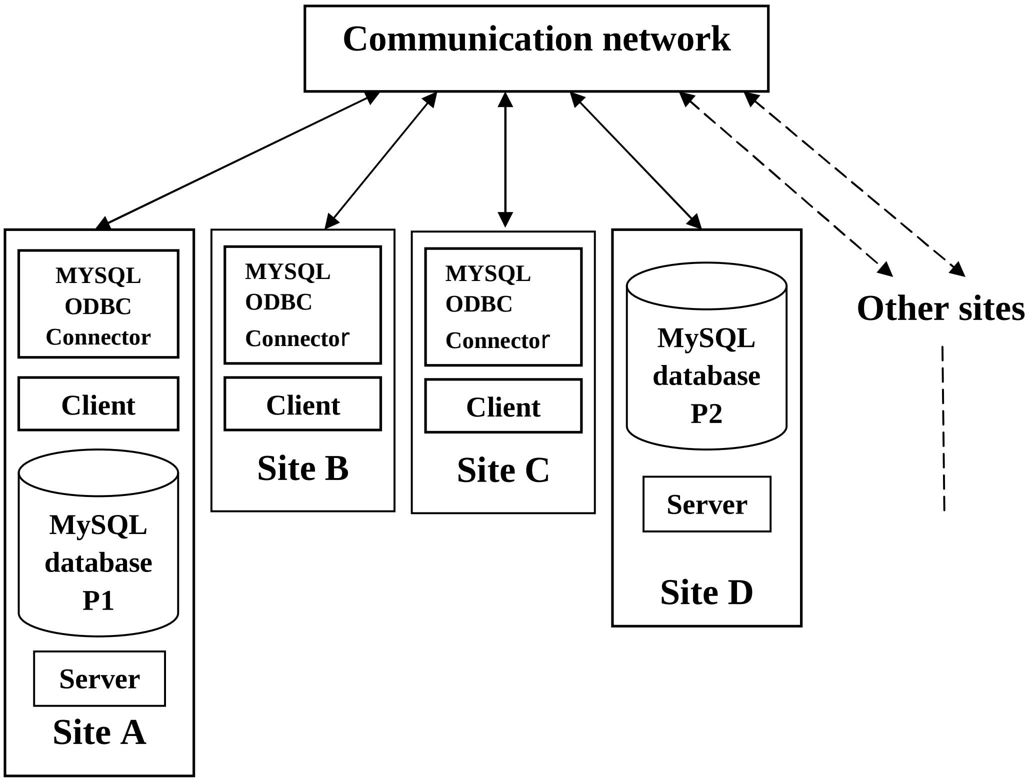


Figure (4.2) PDEL architecture.

4.3 The relational model of PDEL

In the implementation of PDEL, MySQL Server database system used as data storage computer language. The MySQL database is a high-performance relational database management system and it is one of the most widely-used databases in the world. The relations (tables) of PDEL database are:

1. ***Resources:*** this relation is used to save the resources of e-learning. The attributes (columns) of this relation with it's briefly descriptions is shown in Table (4.1).

Table (4.1): Attributes of relation Resources.

Attribute Name	Domain	Size	descriptions	Key type
Cod1	Text	4	Represent the department of this resource.	Composite Primary Key
Cod2	Integer	2	Represent the class that uses this resource as a subject.	
Cod3	Integer	2	The serial number of the subject which use this resource.	
Res_title	Text	50	A name giving to the resource.	
Res_type	Text	8	Includes book, lecture, and research.	
Res_format	Text	8	The physical or digital format of this resource.	
Lang	Text	7	The language of the content of this resource.	
Author	Text	30	The authors of this resource.	
Publisher	Text	20	An entity that published this resource.	
Year	Date		The creation or publication year.	
Location	Text	50	The location of this resource.	
Subject	Text	30	The subject of this resource.	
Kw	Text	30	The keywords of this resource.	
Chk	Boolean	1	Used as a flag in the search operation.	

This relation was divided up into six fragments according to its science these fragments are: comp_res, bio_res, math_res, chem_res, phy_res, and ureq_res. These fragments can be stored at one or more server (location) where it is most frequently used, to reduce the network traffic and make most of the operations are purely local.

2. *Syllabus:* this relation is used to save the syllabus for each class of each department. The attributes of syllabus relation with it's briefly descriptions is shown in Table (4.2).

Table (4.2): Attributes of relation Syllabus.

Attribute Name	Domain	Size	description	Key Type
Dep	Text	1	The department.	Composite Primary Key
Class	Text	1	The class.	
Subject	Text	30	The subject.	
Cod1	Text	4	Represent the department of this resource.	
Cod2	Integer	2	Represent the class that uses this resource as a subject.	
Cod3	Integer	2	The serial number of the subject which use this resource.	

3. Users: this relation is used to save the users information's, it is used as explicit user profile. The attributes (columns) of this relation with it's briefly descriptions is shown in Table (4.3).

Table (4.3): Attributes of relation Users.

Attribute Name	Domain	Size	description	Key Type
User name	Text	30	The name of the user in his account, use to deal with.	Primary Key
Password	Text	14	The password of the user in his account, use to deal with.	
Dep	Text	1	The department of the user.	
Class	Text	1	The class of the user.	

4. Profile: this relation is used to save the previous resource requests (interests) of each user; it is used as implicit user profile. The attributes (columns) of this relation with it's briefly descriptions is shown in Table (4.4).

Table (4.4): Attributes of relation Profile.

Attribute Name	**Domain**	**Size**	**description**	**Key Type**
User name	Text	30	The name of the user in his account, use to deal with.	Foreign Key
Science	Text	1	The science of the previous request.	Composite Primary Key
Subj	Text	30	The subject of the previous request.	
Keyword	Text	30	The keyword of the previous request.	
Ldate	Date		The latest request date.	
Freq	Integer	2	The frequency of this request.	

5. Authors: this relation is used to save the authorize person's information. The attributes (columns) of this relation with it's briefly descriptions is shown in Table (4.5).

Table (4.5): Attributes of relation Authors

Attribute Name	**Domain**	**Size**	**descriptions**	**Key Type**
User name	Text	30	The name of the authorize person in his account, use to deal with.	Primary Key
Password	Text	14	The password of the authorize person in his account, use to deal with.	
Group	Text	4	The group of the authorize person.	

These relations and its fragments are located in to two servers; the first server used to serve two relations: Users and Profile and three fragments of the resources relation; these fragments are: comp_res, math_res, and phy_res. The second server used to serve two relations: Authors and Syllabus and the remaining fragments of the resources relation; these fragments are: bio_res, chem_res, and ureq_res.

4.4 PDEL design

As shown in Figure (4.1) personalization in distributed e-learning consists of many modules so each one of these modules will be describe in detail:

1. User

The user of the e-learning may be manager, authorized person or learner.

2. Manager

The manager is a special type of the authorized persons which has a reserved password, when the user enters as the manager he must insert this password. The manager has full controls over the resources and syllabus and has the permissions to create and update the authorize persons account and determine them controls by classifying them into groups**.** Algorithm (4.1) describes the process of creating new account for the authorize person.

Algorithm (4.1) Create new account for the authorize person

```
Input:
Output/Action: creating new account for the authorize person.
Procedure:
Step1:
    Ask the authorize person to enter his/her information (user name,
    password, confirm password) and the manager give him group.
Step2:
    If user name = "" or password = "" or confirm password = ""
    then report an error message indicating that "Missing
    Information" and go to step2.
```

Continue

```
Step 3:
        If confirm password < > password then clear them, report an
        error message indicating that "Check and Rewrite the Password",
        and go to step2.
Step4:
        Select the user name from the table name (Authors).
        If it is found then clear the contents of the user name and the
        password, report an error message indicating that "This Account
        is already exist", and go to step2.
Step5:
        Insert new record with this information (user name, password,
        and group) into table name (Authors).
```

The manager has the permission to update the account of the authorized person who has an account previously. Algorithm (4.2) describes the process of updating the account of the authorized person.

Algorithm (4.2) Update the authorized person account

```
Input: the authorized person account.
Output/Action: updating the authorize person account.
Procedure:
Step1:
        Ask him about the authorized person account (username and
        password).
Step2:
        If user name = "" or password = "" then report an error message
        indicating that "Missing Information" and go to step2.
Step3:
        Select this account from the table name (Authors).
        If it is not found then report an error message indicating that
        "This Account does not exist" and exit.
Step4:
        Display the information of this account (user name, password,
        and group) to the manager to update it.
```

Continue

```
Step5:
    If the updated (user name = "" or password = "" or  conform password = "") then report an error message indicating that "Missing Information" and go to step5.
Step6:
    If the updated (confirm password < > password) then clear their contents, report an error message indicating that "Check and Rewrite your Password", and go to step5.
Step7:
    Select the new user name from the table name (Authors).
    If it is found then clear the contents of the user name and the password, report an error message indicating that "This Account is already exist", and go to step5.
Step8:
    Update the current record.
```

3. Authorize person

The authorized persons have a control over the resources and syllabus under the permission given to him group from the manager. So they could add, delete, and update the resources and syllabus.

Those persons must have an account number which identifies them and determine them controls to ensure the authentication, they could update them account (algorithm (4.2)) but they couldn't update them group.

When the authorized person wants to use the resources or syllabus then he must login (Deal) with his account to ensure the security. The deal process is described in algorithm (4.3).

Algorithm (4.3) Authorize person deal

```
Input: the account of the authorize person.
Output/Action: deal or not deal.
Procedure:
Step1:
    Dcount = 1.
Step2:
    If Dcount < 4 then ask the authorized person to deal with his/her
    account (username and password).
    Else exit.
Step3:
    If user name = "" or password = "" then report an error message
    indicating that "Missing Information" and go to step2.
Step4:
    • Select this account from the table name (Authors).
    • Dcount = Dcount+1.
    • If it is not found then clear the contents of the user name and
      the password, report an error message indicating that "This
      Account does not exist", and go to step2.
Step5:
    Determine the group of this user name from the table name
    (Authors) to give him control.
Step6:
    Calls add new resource, search about the resource, add syllabus,
    and search about the syllabus.
```

4. Resources

The resources of the e-learning contain books, thesis, researches, and lectures. These resources provides in a variety of formats, which include written materials, CD-ROM, and web links to content on the internet about the subject matter of the course.

The resources relation was divided up into six fragments according to its science these fragments are: comp_res, bio_res, math_res, chem_res, phy_res, and ureq_res, the authorize persons can add, delete, and update an e-learning resources to this fragments restricted on the controls of his group. Algorithm (4.4) describes the process of adding new resource.

Algorithm (4.4) Add new resource

Input: the content of the record fields of table name (resources).
Output/Action: added new resource to the fragment of table resources.
Procedure:
Step1:
    The authorize person enters the content of the record fields (code, resource title, resource type, resource format, language, print date, author, publisher, location, subject, keywords).
Step2:
    Connect to the server that stores the fragment of resources that match the science of this resource.
Step3:
    Select the records that match this resource from the fragment.
    If it is found then report an error message indicating that "This Resource already exists" and exit.
Step4:
    Add this record in to the fragment.

If the authorize persons want to delete, or update the resources he must search about those resources then delete or update them. Algorithm (4.5) describes the process of search about the resource, Algorithm (4.6) describes the process of delete the resource, and Algorithm (4.7) describes the process of update the resource.

Algorithm (4.5) Search about the resource

Input: the fields' content of the search about resources.
Output/Action: The records set that match this search request.
Procedure:
Step1:
Connect to the server that stores the fragment of resources that match the science of the search request.
Step2:
Select the records that match the search request from this fragment.
If there is no records match this request then report an error message indicating that "There are no result" and exit.
Step3:
Call delete or update the resource.

Algorithm (4.6) Delete the resource

Input: the selected record.
Output/Action: deleting this record (resource).
Procedure:
Step1:
Ask the authorize person who want to delete the resource "Are you sure".
If he inserts No then exit.
Step2:
Delete the current record.

Algorithm (4.7) update the resource

Input: the selected record.
Output/Action: update this record.
Procedure:
Step1:
Modify the content of the record fields.
Step2:
Select the records that match the updated resource.
If it is found then report an error message indicating that "This Resource already exists" and exit.
Step3:
Update the table.

5. Syllabus

The syllabus of e-learning contains the syllabus for the departments of the college of science.

The authorize persons can add, delete, and update the syllabus restricted in the controls of his group. Algorithm (4.8) describes the process of adding the syllabus.

Algorithm (4.8) Add syllabus

```
Input: the content of the record fields of table name (syllabus).
Output/Action: added new record to the table name (syllabus).
Procedure:
Step1:
    The authorize person enter the content of the record fields
    (department name, class, subject).
Step2:
    Select this record from the table name (syllabus).
    If it is found then report an error message indicating that "This
    Syllabus already exists" and exit.
Step3:
    Add this record in to the table name (syllabus).
```

If the authorize persons want to delete, or update the syllabus at first he must search about it then delete or update the syllabus. Algorithm (4.9) describes the process of search about the syllabus; Algorithm (4.10) describes the process of delete from the syllabus; Algorithm (4.11) describes the process of update the syllabus.

Algorithm (4.9) Search about the syllabus

Input: the field's content of the search about the syllabus.
Output/Action: The records set that match this search request.
Procedure:
Step1:

Select the records that match the search request in the table name (syllabus).

If there is no any record match this request then report an error message indicating that "There are no result" and exit.

Step2:

Call delete or update syllabus.

Algorithm (4.10) Delete from the syllabus

Input: the selected record.
Output/Action: deleting this record (subject) from the table name (syllabus).
Procedure:
Step1:

Ask the authorize person who want to delete from the syllabus "Are you sure?".

If he inserts o then exit.

Step2:

Delete the current record.

Algorithm (4.11) update the syllabus

Input: the selected record.
Output/Action: update this record.
Procedure:
Step1:

Modify the content of the record fields.

Step2:

Select this updated record from the table name (syllabus).

If it is found then report an error message indicating that "This Syllabus already exists" and exit.

Step3:

Update the table.

6. Learner

The learner use PDEL to search about resources that match his/her needs, it offers to him two ways of searching these are *search* and *personal search*.

At first, the learner must be identifying himself to ensure the authentication by creating new account to get the permission of searching and create his explicit user profile. Algorithm (4.12) describes the process of creating new account for the learner.

Algorithm (4.12) Create new account for the learner

Input: the learner's information.
Output/Action: creating new account for the learner.
Procedure:

Step1:
Ask the learner to enter his/her information (user name, password, confirm password, department, and class).

Step2:
If user name = "" or password = "" or conform password = "" then report an error message indicating that "Missing Information" and go to step1.

Step 3:
If confirm password < > password then clear the content of them, report an error message indicating that "Check and Rewrite your Password", and go to step1.

Step4:
If class < > "" and department = "" then report an error message indicating that "Enter your department" and go to step1.

Step5:
Select the user name from the table name (users).
If it is found then report an error message indicating that "This Account is already exist" and exit.

Step6:
Insert new record with this information (user name, password, department, class) in to the table name (users) and this is used as an explicit profiling.

The learner who has an account number can update his/her account (the explicit user profile information) this process is described in algorithm (4.13).

Algorithm (4.13) Update the learners account

Input: the learners account.
Output/Action: updating the learners account information.
Procedure:

Step1:
Ask the learner to enter his/her account (username and password).

Step2:
If user name = "" or password = "" then report an error message indicating that "Missing Information" and go to step1.

Step3:
Select this account from the table name (users).
If it is not found then report an error message indicating that "This Account does not exist" and exit.

Step4:
Display the information of this account (user name, password, department, class) to the user to update it.

Step5:
If the new (user name = "" or password = "" or confirm password = "") then report an error message indicating that "Missing Information" and go to step4.

Step6:
If the new (confirm password < > password) then clear the content of them, report an error message indicating that "Check and Rewrite your Password", and go to step4.

Step7:
If the new (class < > "" and department = "") then report an error message indicating that "Enter your department" and go to step4.

Step8:
If the new user name < > user name then

- Select the new user name from the table name (users).
- If this user name found then report an error message indicating that "This Account is already exist" and exit.

Step9:
Update the current record.

When the learner begin his search, he must login (deal) with his account to ensure the security. The deal process is described in algorithm (4.14).

Algorithm (4.14) Deal

```
Input: the learner's account.
Output/Action: deal or not deal.
Procedure:
Step1:
    Dcount = 1.
Step2:
    If Dcount < 4 then ask the learner to deal with his/her account
    (user name and password).
    Else exit.
Step3:
    If user name = "" or password = "" then report an error message
    indicating that "Missing Information" and go to step2.
Step4:
    • Select this account from the table name (users).
    • Dcount = Dcount+1.
    • If it is not found then clear the contents of the user name and
      the password, report an error message indicating that" This
      Account does not exist", and go to step2.
```

7. Search

The learner can search using one of the two search methods (search or personal search). If the learner selects the *search* then he can request any resources and this request (current request) will be saved in the implicit user profile of this learner (the table name Profile) as an interest (previous resource request) with the latest request date (Ldate) and the frequency of this request (Freq). If this learner hasn’t implicit user profile, it must be created in the search process to save the current request. Algorithm (4.15) describes the search process.

Algorithm (4.15) Search

Input: the learner's request.

Output/Action: the records set that match this request.

Procedure:

```
Step1:
    Connect to the server that stores the fragment of resources that
    match the science of the learner request.
Step2:
    Select the records that match the current request from this
    fragment and display these records to the learner.
    If there is no record match this request then report an error
    message indicating that "There are no result" and exit.
Step2:
    Select the records that match the current user name from the
    table name (Profile).
Step3:
    Found= False
    While (Not EOF) Do
        If request of the current record = current request Then
            • Ldate = Now
            • Freq = Freq + 1
            • Found = True
            • Exit Do
        End If
        Move to next record
    End While
Step4:
    If Found = False Then
        • Insert the current request in the table name (Profile) with
          username = current username
        • Ldate = Now
        • Freq = 1
    End If
```

8. Personal search

If the learner would like to search the resources that he requested them previously or search some resources that related to his syllabus then the *personal search* is more suitable search method for him because it make the search about this request fast and very easy. The learner has three ways of personal search and he must choose one of them, these are personal search about previous resource request according to the latest date, personal search about previous resource request according to the frequency of this request, and personal search about syllabus.

9. Personal search about previous resource requests according to the latest date

When the learner selects this search method and enters the number of requests that he wants to search then it returns the resources about these requests which has latest request date (more interest to him), and saved in the implicit user profile as the previous resource request (interest) from the search process. Algorithm (4.16) describes the process of Personal search of the previous resource requests according to the latest date.

Algorithm (4.16) Personal search about previous resource requests according to the latest date

Input: the user name and the number of latest previous requests.
Output/Action: the more interest resources according to the latest date.
Procedure:
Step1:
Select the records that match the user name from the table name (Profile) ordered by the Ldate descending. The number of records = RecNum, this records used as an interest requests.
Step2:
If RecNum = 0 then send a message to the user telling him that "You haven't any previous search" and exit.
Step3:
Ask the learner to enter the number of requests (Num).

Continue

```
Step4:
      If RecNum< Num then Num=RecNum.
Step5:
      Create table (temporary table) of the name = current user name
      with the same fields of the table resources.
Step6:
      Move to the first request from the set of interest request that
      selected from the table name (Profile), and this is used as the
      current request.
      RecNum=0
      While (Not EOF) OR (RecNum< Num) Do
          • Connect to the server that stores the fragment of resources
            that match the science of the current request.
          • Select the records that match the current request from this
            fragment of resources and add these records to the
            temporary table.
          • RecNum=RecNum+1
          • Move to the next request from the set of interest requests
            that selected from the table name (Profile) and this is used
            as the current request.
      End While
Step7:
      Select the records from the temporary table and display these
      records to the user.
Step8:
      Drop table (temporary table) of the name = current user name if
      exist.
```

10. Personal search about previous resource requests according to their frequency

When the learner selects this search method and enters the number of requests that he wants to search then it returns the resources of the more selected requests, these requests has biggest frequency, and were saved in the implicit user profile as the previous resource request (interest) from the search process. Algorithm (4.17) describes the process of Personal search about previous resource requests depending on the frequency.

Algorithm (4.17) Personal search about previous resource requests according to the frequency

Input: the user name and the number of latest previous requests.

Output/Action: the more interest resources according to the frequency.

Procedure:

```
Step1:
    Select the records that match the user name from the table name
    (Profile) ordered by the Freq descending, the number of
    records=RecNum, and this records used as an interest requests.
Step2:
    If RecNum = 0 then send a message to the user telling him that
    "You haven't any previous search" and exit.
Step3:
    Ask the learner to enter the number of requests (Num).
Step4:
    If RecNum< Num then Num=RecNum.
Step5:
    Create table (temporary table) of the name = current user name
    with the same fields of the table resources.
Step6:
    Move to the first request from the set of interest request that
    selected from the table name (Profile), and this is used as the
    current request.
    RecNum=0
    While (Not EOF) OR (RecNum<Num) Do
        • Connect to the server that stores the fragment of resources
          that match the science of the current request.
        • Select the records that match the current request from this
          fragment of resources and add these records to the
          temporary table.
        • RecNum=RecNum+1
        • Move to the next request from the set of interest requests
          that selected from the table name (Profile) and this is used
          as the current request.
    End While
```

Continue

```
Step7:
      Select the records from the temporary table and display these
      records to the user.
Step8:
      Drop table (temporary table) of the name = current user name if
      exists
```

11. Personal search about syllabus

When the learner selects this search method then it returns all the subjects of his syllabus, if he selects any one of those subjects then all the resources related to that subject will be returned. Algorithm (4.18) describes the process of personal search about syllabus.

Algorithm (4.18) Personal search about syllabus

```
Input: the user name.
Output/Action: the records set that match the selected subject.
Procedure:
Step1:
      Select class and department of the current user from the table
      name (users).
Step2:
      If class = "" Then report an error message indicating that "You
      are not a student" and exit.
Step3:
      Select the subjects related to this class and department from the
      table (syllabus).
Step4:
      Display the subjects of the user's syllabus.
Step5:
      If the user selects any one of those subjects to find the resources
      related to it, then a connection to the server that stores the
      required subject will be established, and select the resources of
      this subject from this fragment.
```

4.5 PDEL implementation interfaces

The Main form contains three buttons as shown in Figure (4.3):

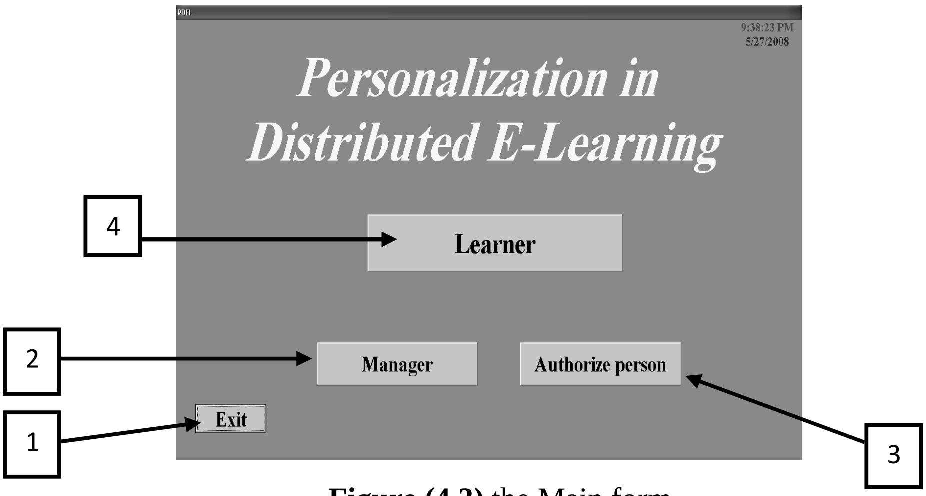


Figure (4.3) the Main form.

1. ***Exit:*** ends the program execution.
2. ***Manager:*** at first he must insert the manager password then the Manager account no. frame will be displayed in the Main form as shown in Figure (4.4).

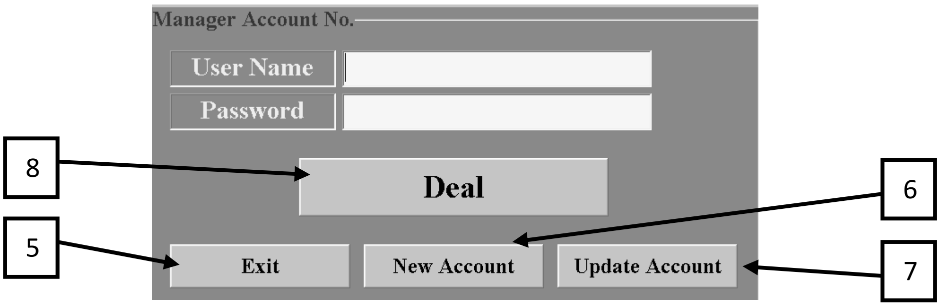


Figure (4.4) the Manager account no. frame.

3. ***Authorize person:*** the Authorize person account no. frame will be displayed in the Main form as shown in Figure (4.5).

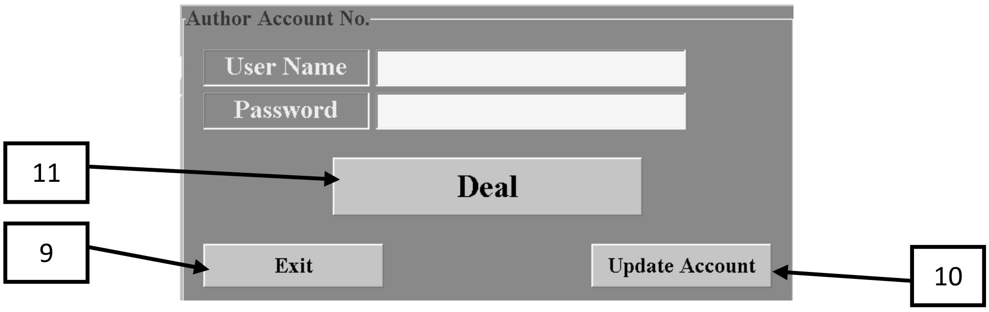


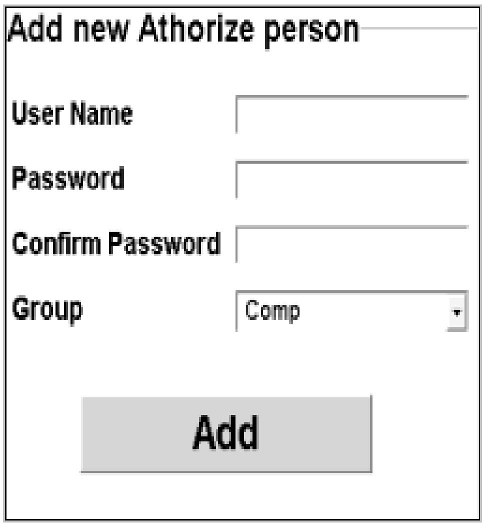
Figure (4.5) the Authorize person account no. frame.

4. ***Learner:*** the Learner account no. frame will be displayed in the Main form as shown in Figure (4.17).

The Manager account no. frame in Figure (4.4) contains four buttons:

5. ***Exit:*** close the Manager account no. frame.

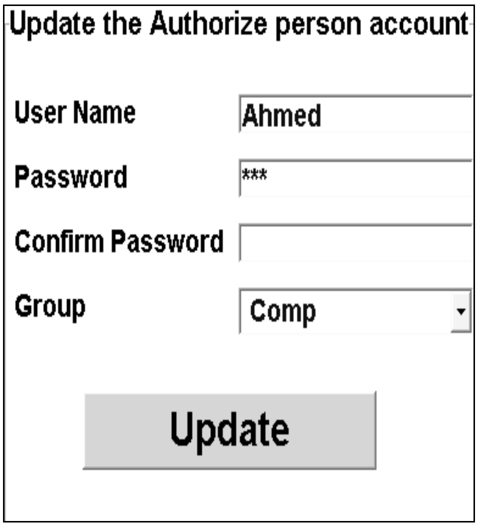
6. ***New Account:*** open Add new authorize person frame as shown in Figure (4.6), when he fill the fields of this frame. Then this user will be identified as an authorize person (has the authorize account no.).



Group Comp

Figure (4.6) Add new authorize person frame.

7. Update Account: ask the manager to enter the account no. of the authorized person who want to update his account then open Update the authorize person account frame which is filled with the information of this account as shown in Figure (4.7), when he update this fields and click on update button then the authorized person account will be updated.



Update the Authorize person account

User Name: Ahmed

Password: ***

Confirm Password:

Group: Comp

Figure (4.7) Update the authorize person account frame.

8. Deal: when the manger enter the username and the password of the authorized person account and click on Deal button in Figure (4.4) then the Authorize person form was opened (Figure (4.8)).

The Authorize Person account no. frame in Figure (4.5) contains four buttons:

9. Exit: close the Authorize person account no. frame.

10. Update Account: ask the authorize person to enter his account then open Update the authorize person account frame which contains the user name and the password of this account without the group because the manager only has a permission to update the group of the authorized person when he update this fields and click on update button then the authorized person account will be updated.

11. ***Deal:*** when the authorize person enter his username and password click on Deal button in Figure (4.5) then the Authorize person form was opened (Figure (4.8)).

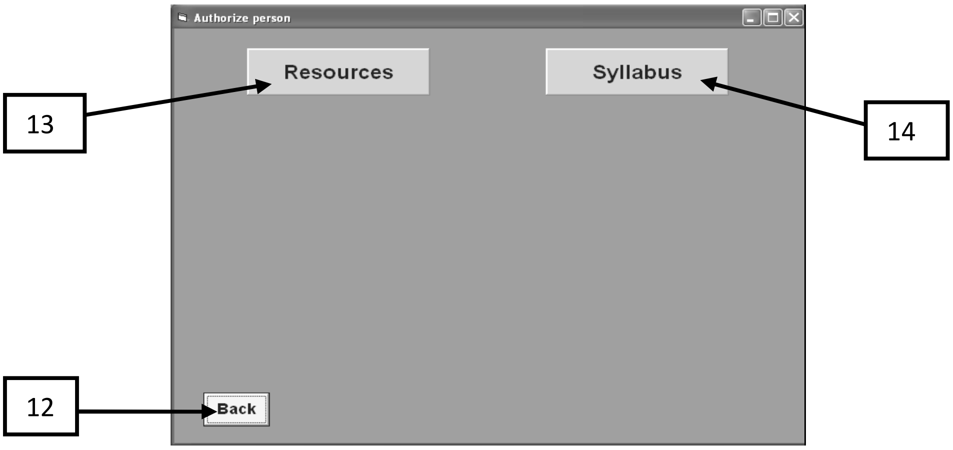


Figure (4.8) the Authorize person form.

As shown in Figure (4.8), the Authorized person form contains three buttons:

12. ***Back:*** Return to Main form.
13. ***Resources:*** the resources which this authorize person have a control over them will be displayed in the Authorize person form as an active buttons, this buttons will be displayed depending on the group of this authorize person (the permission giving for him). For example if the authorize person was in computer group then he has a control over computer resources only, so that the computer button displayed in an active form as shown in Figure (4.9), if he choose this button then the Resources form will be opened. This form is shown in Figure (4.11).

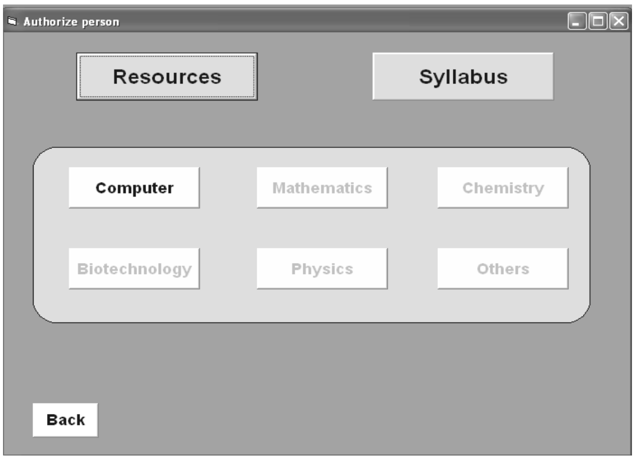


Figure (4.9) The Authorize person form with resources.

14. Syllabus: the departments syllabus which this authorize person have a control over them will be displayed in the authorize person form as an active buttons, this buttons will be displayed depending on the group of this authorize person. If the same authorized person which choose the resources press on the syllabus button, then he has the permission over the computer department syllabus only and the computer button will be displayed in an active form as shown in Figure (4.10), if he choose this button then the Syllabus form will be opened. This form is shown in Figure (4.14).

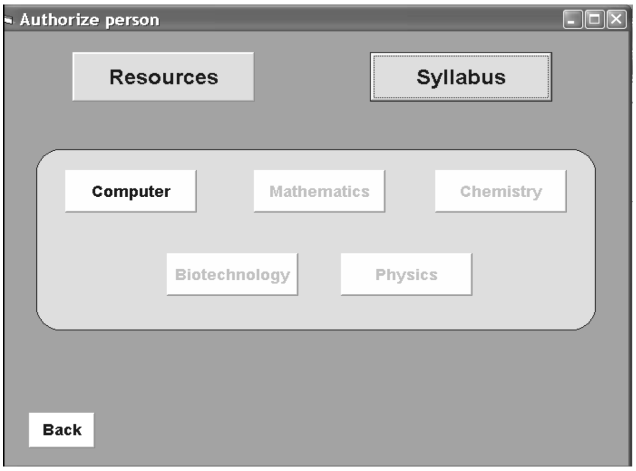
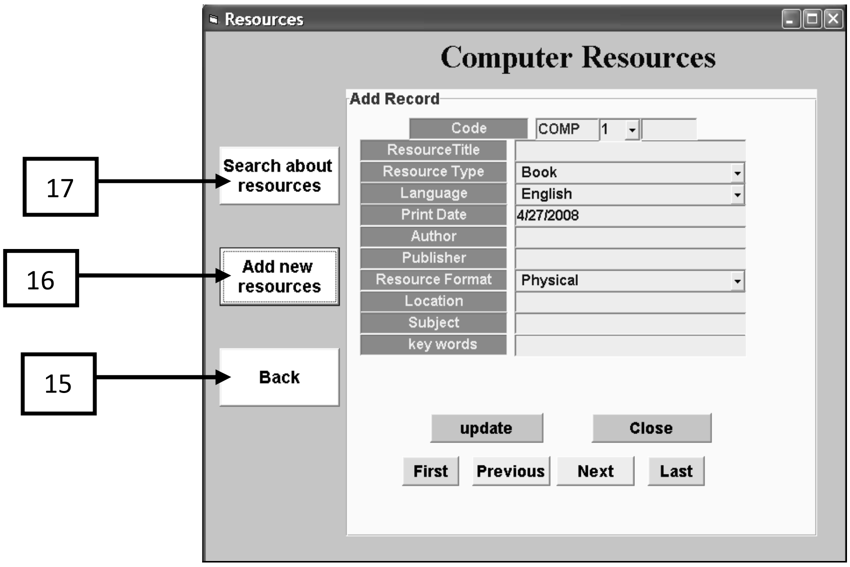


Figure (4.10) the Authorize person form with syllabus.

In Figure (4.11) the Resources form contains three buttons:

15. Back: Return to the Authorize person form.

16. Add new resources: open Add record frame, when the authorize person fill this fields and choose the update button then this resource will be inserted in to the resources table. The First, Previous, Next, and Last buttons in Add record frame used to exchanged between the records, and Close button used to close the frame.



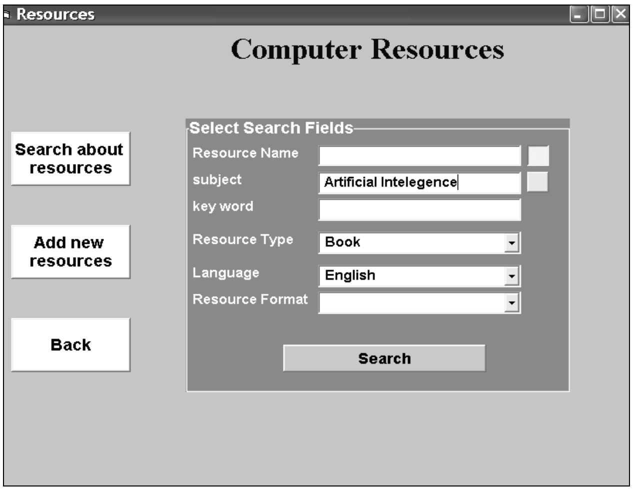
Computer Resources

Add Record

Code	COMP 1
ResourceTitle	
Resource Type	Book
Language	English
Print Date	4/27/2008
Author	
Publisher	
Resource Format	Physical
Location	
Subject	
key words	

Figure (4.11) the Resources form with add record frame.

17. Search about resources: open Select search fields frame as shown in Figure (4.12). When the authorize person fill the selected fields and press on the Search button then this frame will be closed and the Search resources frame will be opened as shown in Figure (4.13).

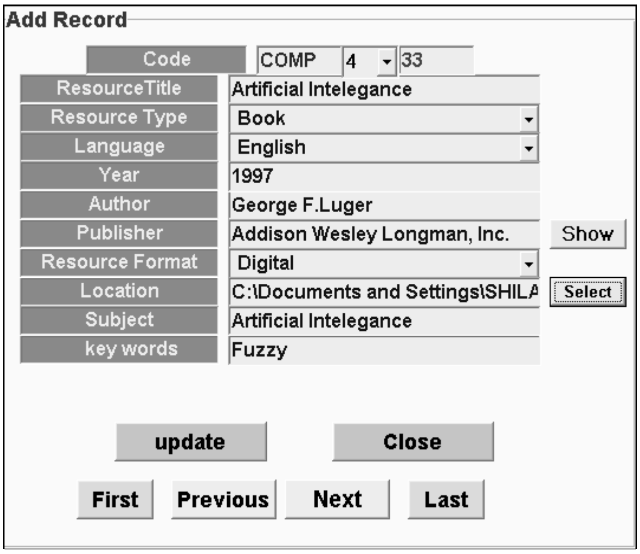


Computer Resources

Select Search Fields

Resource Name:
subject: Artificial Intelegence
key word:
Resource Type: Book
Language: English
Resource Format:

Figure (4.12) the Resources form with Select search fields frame.



Add Record

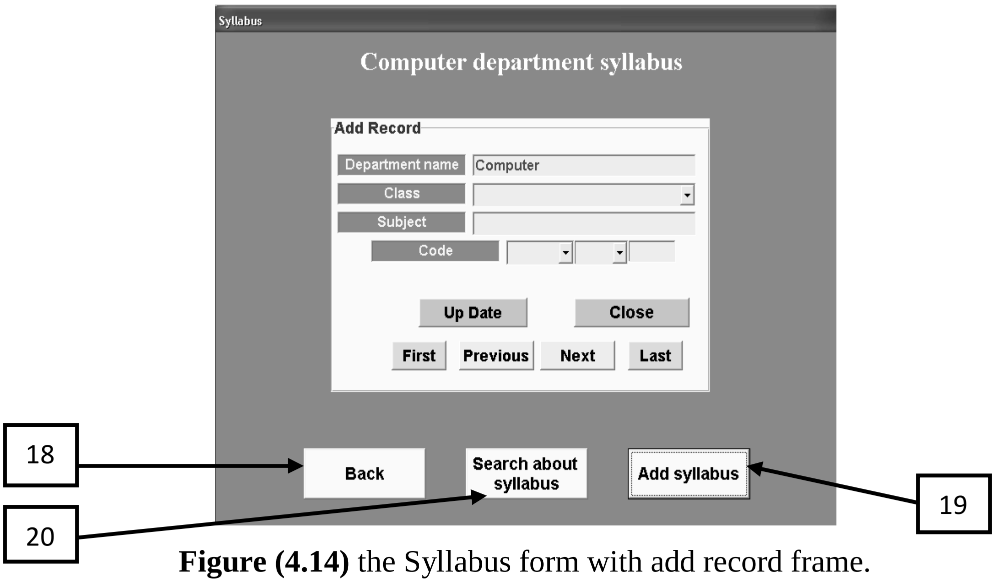
Code	COMP 4 33
ResourceTitle	Artificial Intelegance
Resource Type	Book
Language	English
Year	1997
Author	George F.Luger
Publisher	Addison Wesley Longman, Inc.
Resource Format	Digital
Location	C:\Documents and Settings\SHILA
Subject	Artificial Intelegance
key words	Fuzzy

Figure (4.13) the Search resources frame.

The Search resources frame contains the full information of all resources that match the search fields. It contains nine buttons these are:

- ***Select:*** Used to select the path of the resources if the resource of digital format and it is saved in the server site.
- ***Show:*** used to open and display the resource if it is of digital format and saved in the server site.
- ***Delete:*** used to delete the current record.

- ***Modify:*** used to modify the information of the current record.
- ***First, Previous, Next, and Last:*** used to exchange between those records.
- ***Close:*** used to close this frame.


Syllabus

Computer department syllabus

Add Record

Department name: Computer

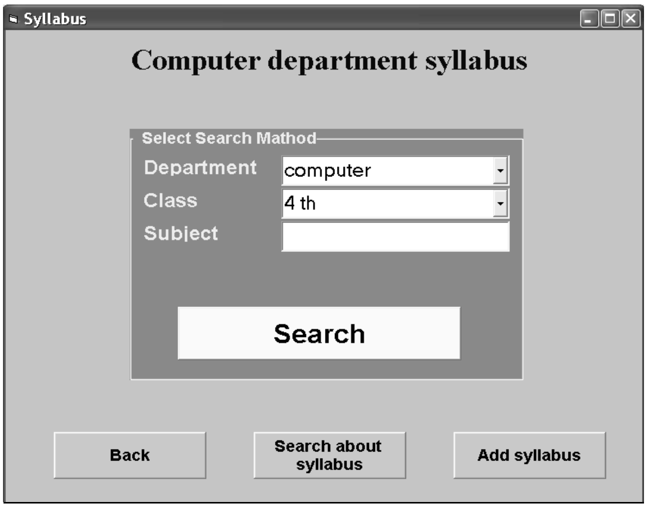
Figure (4.14) the Syllabus form with add record frame.

In figure (4.14) the Syllabus form contains three buttons:

18. Back: return to the Authorize person form.

19. Add syllabus: open Add record frame, when the authorize person fill this fields and choose the update button then this resource will be inserted into the Syllabus table. The First, Previous, Next, and Last buttons in Add record frame used to exchanged between the records, and Close button used to close the frame.

20. Search about syllabus: open Select search fields frame in figure (4.15). When the authorize person fill the selected fields and choose the search button then this frame will be closed and open the Search syllabus frame as shown in figure (4.16).



Syllabus

Computer department syllabus

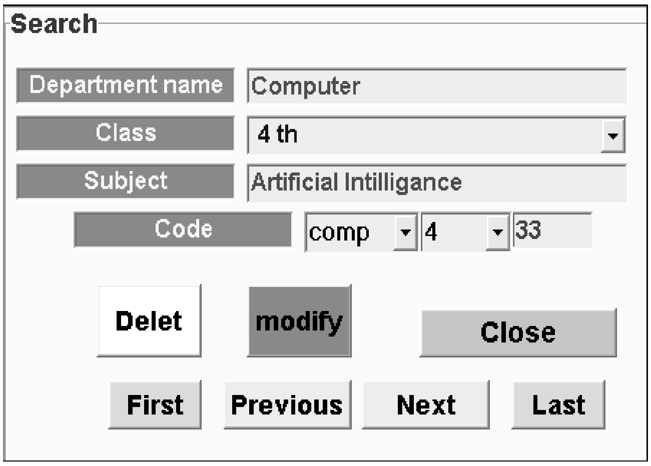
Select Search Mathod

Department: computer

Class: 4 th

Subject:

Figure (4.15) the Select search fields frame.



Search

Department name: Computer

Class: 4 th

Subject: Artificial Intilligance

Code: comp 4 33

Figure (4.16) the Search Syllabus frame.

As shown in figure (4.16) the Search syllabus frame contains the information of the Syllabus that match the search fields in figure (4.15). It contains seven buttons these are:

- ***Delete:*** used to delete the current record.
- ***Modify:*** used to modify the information of the current record.
- ***First, Previous, Next, and Last:*** used to exchange between those records.
- ***Close:*** used to close this frame.

If the user entered as learner, then the Learner account no. frame (Figure (4.17)) will be displayed which contains four buttons:

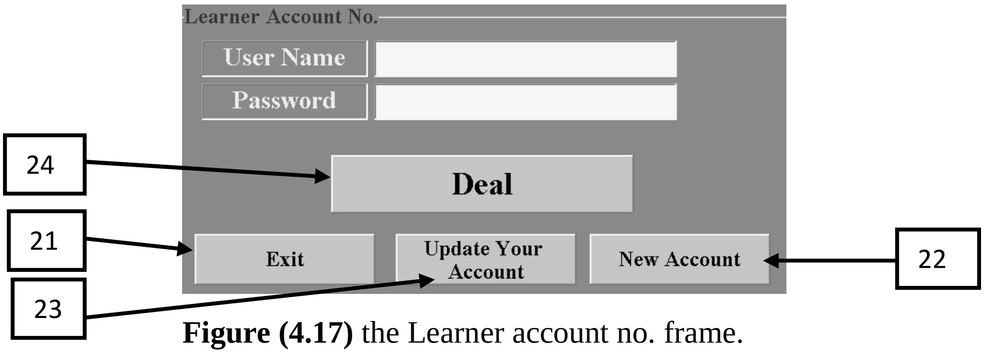
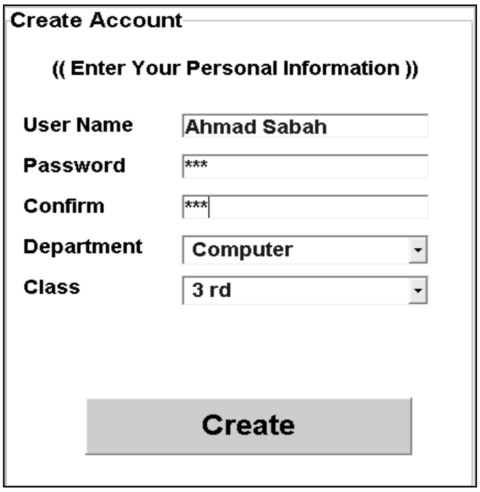


Figure (4.17) the Learner account no. frame.

21. Exit: close the Learner account no. frame.

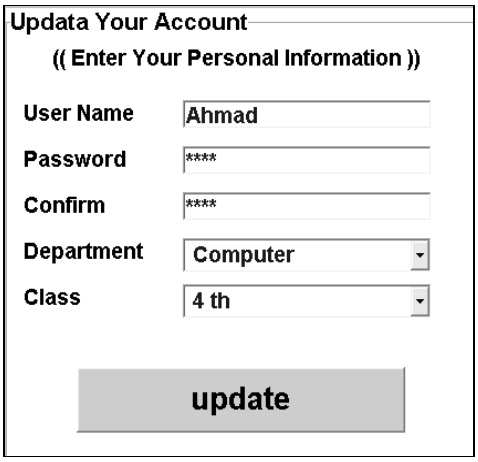
22. New Account: open Create account frame as shown in figure (4.18), when the new user fill the fields of this frame then he will identified as a learner (has the learner's account no. and user profile).



User Name: Ahmad Sabah
Password: ***
Confirm: ***
Department: Computer
Class: 3 rd

Figure (4.18) Create account frame.

23. Update Account: ask the learner to enter his/her account no. which he wants to update it then open Update your account frame which is filled with the information of this account as shown in figure (4.19), when he update these fields and click on update then the learner account will be updated.



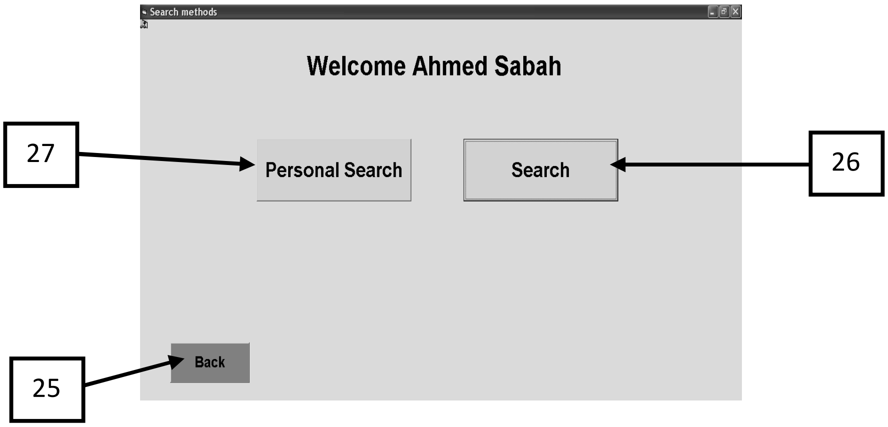
Updata Your Account

((Enter Your Personal Information))

User Name	Ahmad
Password	****
Confirm	****
Department	Computer
Class	4 th

Figure (4.19) Update your account frame.

24. Deal: when the learner enter the user name and the password of his account in the Learner account no. frame and click on Deal button in figure (4.17) then the Search methods form will be opened as shown in figure (4.20).



Search methods

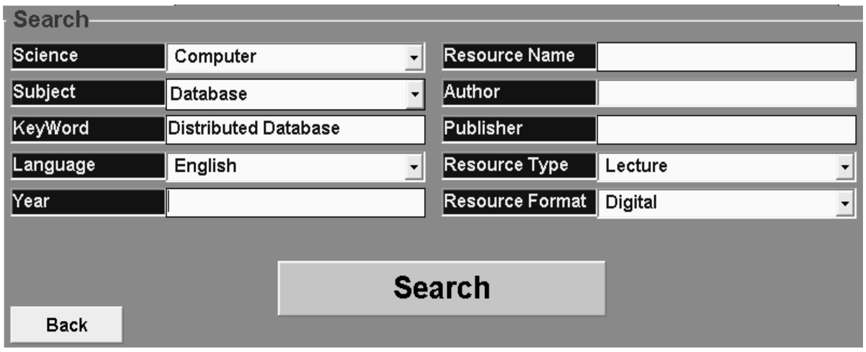
Welcome Ahmed Sabah

Figure (4.20) the Search methods form.

The Search methods form contains three buttons:

25. Back: return to the Main form.

26. Search: open the Search frame as shown in figure (4.21).When the learner fills the fields that he wants to search about them and press on the search button then all the resources that match his request will be displayed in the Resources frame as shown in figure (4.23). If the Learner presses on Back button then he will return to the Search methods form.

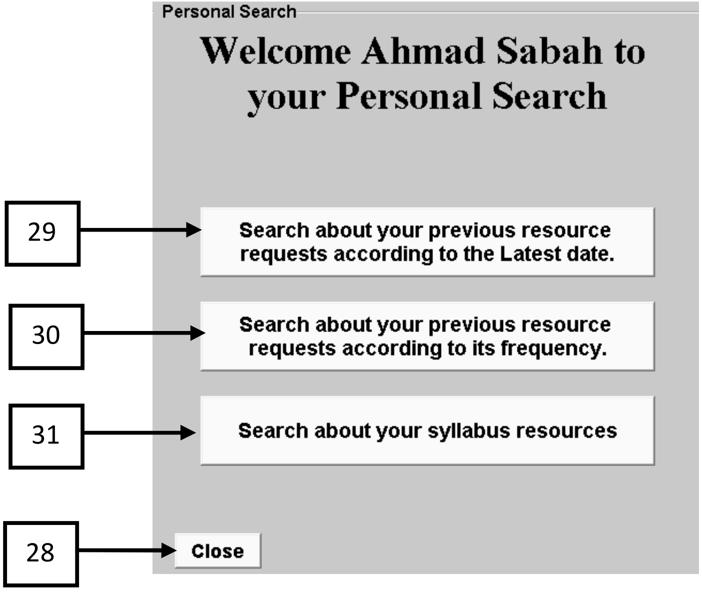


Search

Science	Computer	Resource Name	
Subject	Database	Author	
KeyWord	Distributed Database	Publisher	
Language	English	Resource Type	Lecture
Year		Resource Format	Digital

Figure (4.21) the Search frame.

27. Personal Search: open the Personal search frame which contains four buttons as shown in figure (4.22).



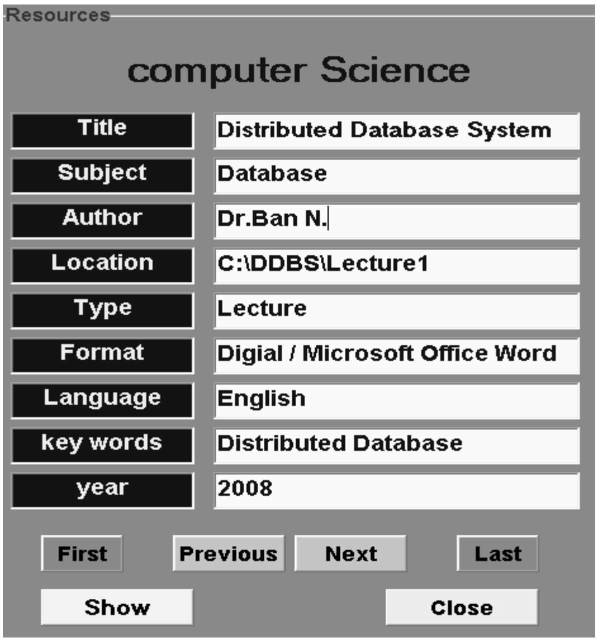
Personal Search

Welcome Ahmad Sabah to your Personal Search

29
Search about your previous resource requests according to the Latest date.
30
Search about your previous resource requests according to its frequency.
31
Search about your syllabus resources
28
Close

Figure (4.22) the Personal search frame.

28. ***Close:*** close the Personal search frame and return to the Search methods form.
29. ***Search about previous resource requests according to the latest date:*** if the learner selects this button then he must enter the number of requests that he wants to search about, all the resources that match his previous resource requests according to the latest date will be displayed automatically in the Resources frame as shown in Figure (4.23).
30. ***Search about previous resource requests according to its frequency:*** if the learner selects this button then he must enter the number of requests that he wants to search about, all the resources that match his previous resource requests according to its frequency will be displayed automatically in the Resources frame as shown in Figure (4.23).
31. ***Search about syllabus:*** if the learner selects this button then all the subjects of his syllabus will be displayed in the Subjects frame as shown in Figure (4.24), then if he selects any one of those subjects, all the resources related to that subject will be displayed automatically in the Resources frame in Figure (4.23).

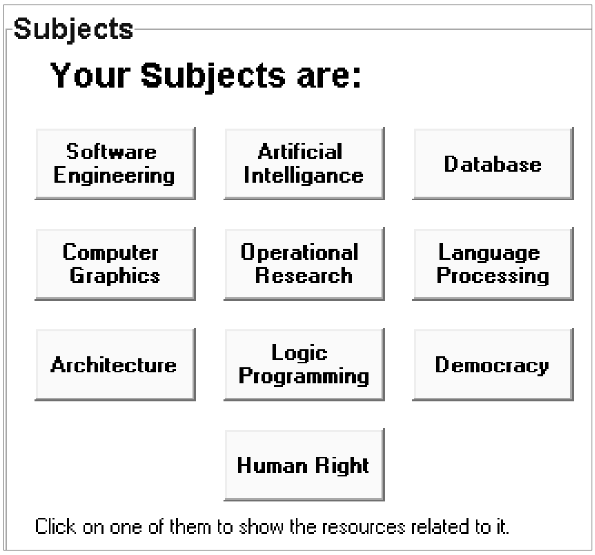


Resources

computer Science

Title	Distributed Database System
Subject	Database
Author	Dr.Ban N.
Location	C:\DDBS\Lecture1
Type	Lecture
Format	Digial / Microsoft Office Word
Language	English
key words	Distributed Database
year	2008

Figure (4.23) the Resources frame.



Subjects

Your Subjects are:

Click on one of them to show the resources related to it.

Figure (4.24) the Subjects frame.

Chapter One

General Introduction

1.1 Distributed database

A database is a collection of data that is stored and maintained at one central location. A database is controlled by a database management system. The user interacts with the database management system in order to utilize the database and transform data into information [Mit05].

The developments in computer networking technology and database system technology resulted in the development of distributed databases in the mid 1970s. It was felt that many applications would be distributed in the future and therefore the databases had to be distributed also. A distributed database system includes a distributed database (DDB), a distributed database management system (DDBMS) and a network for interconnection [Kos02].

A distributed database is a collection of data that belong logically to the same system but are spread over the sites of a computer network. This definition emphasizes two equally importance aspects of a distributed database:

a. Distribution: the data are not resident at the same site, so that DDB can be distinguished from a single centralized database.

b. Logical correlation: the data have some properties which tie them together, so that DDB can be distinguished from a set of local databases or files which are resident at different sites of a computer network.

A distributed database management system is the software system that supports the creation and maintenance of distributed database. [Cer85]

A distributed database system consists of a collection of sites, connected together via a communication network, as shown in Figure (1.1). The main parts in this figure are [Zho01]:

a. Client (C): the front end of a DDBS where the access requests are issued.

b. Server (S): the backend of a DDBS where the database is stored.

c. Communication system (CS): it enables the communication between clients and servers.

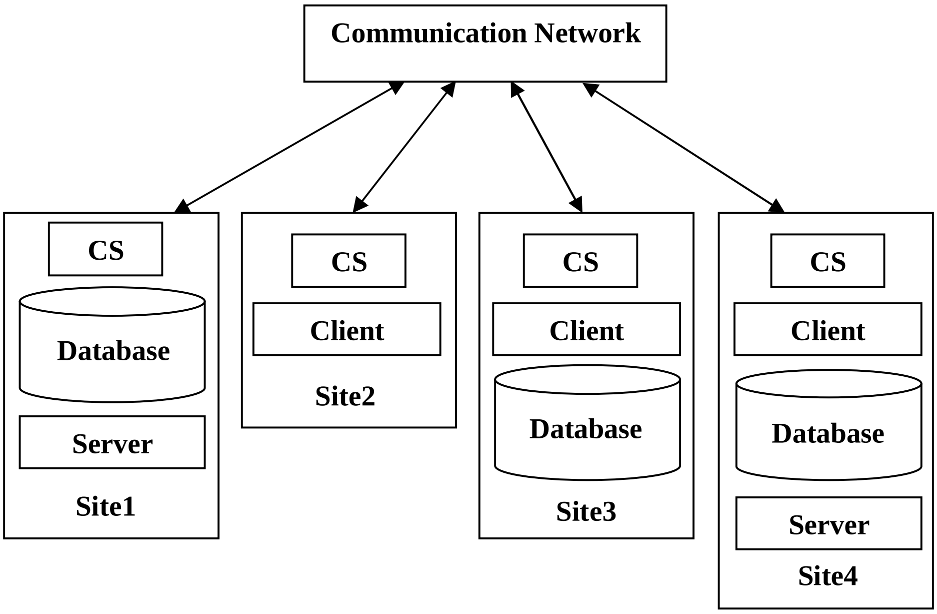


Figure (1.1) Architecture of DDBS

1.2 Personalization

Personalized information systems aims at giving the individual user optimal support in accessing, retrieving, and storing information. The individual requirements of the user are to be taken into account in such different dimensions like the current task, the goal of the user, the context in which the user is requesting the information, the previous information requests or interactions, the working process s/he is involved in, the level of expertise, the device s/he is using to display the information, the bandwidth and availability of the communication channel, the abilities (disabilities or handicaps) of the user, his/her time constraints [Bal05].

1.3 E-Learning

E-learning in general can be seen as technology-delivered or technology-enhanced learning. In one scenario of usage the learners and the instructors can be physically separated (can never or rarely meet for face-to-face lectures, discussions, etc.) and thus all the process of studying and teaching is technology-mediated. In another scenario the traditional learning approaches (e.g. university education) can be supported with complementary services, like online delivery of the learning materials, syllabus, schedules and etc. The goals of the different e-learning systems can differ and so the functionalities offered by them – an industrial company can have as a goal to support the continuous delivery of up-to date knowledge to its employees while the goal of a university institution is to deliver guided education in particular area and develop the students thinking [Tri03].

1.4 Literature Survey

Some of other related works are listed below:

1."Recommender system for E_Commerce Data" [Bas04].

In this MSc. thesis, a recommender system is built that uses different recommendation methods to advice a customer the best items that suit his/her interest from a selected site. In a typical recommender system people provide recommendations as inputs which the system then aggregates and direct to appropriate recipients. Method accuracy depends on several factors, these factors are: the speed of that method, the accuracy which is different from customer to another and the psychological conditions of the recommender system users.

2."Personalization in Distributed e-Learning Environments" [Dol04].

This paper shows how to realize personalized learning support in distributed learning environments based on Semantic Web technologies. The authors described an approach to bring personalization to the semantic web for the area of education and learning. They have shown how personalization functionalities can be embedded into semantic web services, supported by other services for retrieving learning resources or user information.

3."A Study of User XQuery Method for Personalization Recommendation in Semantic Web" [Kim05].

This work is used to alleviate the problem of the information overload when browsing and searching the World Wide Web, because the number of web pages increases dramatically, personalization becomes a popular remedy to customize the Web environment towards a user's preference. The authors presented a User XQuery method for personalization recommendation in semantic web that used in the domain of e-commerce and m-commerce based ubiquitous environment. Also, they designed personalization framework in ubiquitous environment based on both e-commerce and m-commerce and presented the interaction of user profile including user keyword and XQuery pattern in semantic web.

4."Design and Implementation of Web Course Tools (WebCT)" [Abb05]

In this MSc. thesis, WebCT is a web based instruction tool that enables instructors to create and customize their course for distance higher education. The WebCT is presented as a powerful tool that facilitates the creation of sophisticated web based educational environments such as registering, studying, course, content, evaluation, communication and exam tools. The security and privacy of general web based learning systems are evaluated, by adding the security and privacy requirements.

5. "Techniques for Personalization in E-learning" [Abd06].

In this MSc. thesis, the author is going to use web usage mining technique for creating an adaptive recommendation system for navigating a web site. This system can be used as possible approach in the e-learning domain to improve learner's satisfaction, by attempting to personalize the learning navigation environment for the learner.

6."Customization and personalization in the Internet economy" [Reu06].

This MSc. thesis support customization and personalization for pure digital products in the internet economy to dramatically decrease search costs for consumers. This thesis builds upon theories about digital products, mass customization and variety. The main motivations for this thesis are the specific characteristics of digital products and the internet economy. The characteristics of digital products enable the customization of digital products, and in combination with the internet, these characteristics make it possible for consumers to have virtually unlimited choice.

1.5 Aim of Thesis

This research aims to provide the personalized access to learning resources in a distributed e-learning database. The proposed DDBMS architecture is Two-Tier Multiple-Client/Multiple-server model which provides fragmentation transparency, and location transparency.

Personalization meets the user's personal needs, making the access to resources faster and easier. The personal needs are to be taken into account in different ways like the current request, the previous resources requests or the education level of the user (syllabus).

1.6 Thesis outlines

This is the summary of the contents of the subsequent chapters of this thesis:

- **Chapter two:** this chapter presents the concepts need in distributed database system at which introduction to DDBS, a description of database systems, advantages and disadvantages of DDBS, structure of DDB, and distributed DBMSs.
- **Chapter three:** this chapter describes the e-learning, personalization and personalization in distributed e-learning.
- **Chapter four:** this chapter presents the proposed system architecture, the relational model that used in this system, the algorithms that used to implement this system, and the implementation interfaces.
- **Chapter five:** this chapter explores conclusions of this work, and the suggestions for future works.

Chapter Three

Personalization in the E-Learning

3.1 Electronic Learning (E-Learning)

E-Learning is any learning that takes place using a network for the delivery, interaction, or facilitation of learning. It is most commonly associated with activities. The delivery and management of training or learning uses computer and technology. Advantages are its flexibility and accesses to any number of courses make it particularly attractive to corporate trainers. Educational uses of e-learning have grown substantially. Students are setting their own time schedules, learning from each other and from instructors all over the globe. Most higher education institutions offer e-learning opportunities. It is recognized that E-learning is becoming a significant force in education and training. The number of students involved grows and the demand will probably exceed the availability of course material. [Tay02]

Other common names for e-learning include Computer Based Learning (CBL), Web Based Learning (WBL) and Multimedia Based Learning (MBL) [Min03]. E-learning takes place in interaction between instructors and students it can occur in three formats [Che06]:

1. Synchronous e-learning: instructions that is led by a facilitator in real time. Example of synchronous interactions includes conference calls, instant-messaging, video conferences, whiteboard sessions and sessions in online classrooms.

2. Asynchronous e-learning: instructions that is not constrained by geography or time. Every one involved in an asynchronous activity performs his or her part on his or her own time.
3. Some combination of both synchronous and asynchronous e-learning.

E-learning includes instruction delivered via all electronic media including the internet, intranets, satellite broadcasts, audio/video tape, interactive TV, and CD-ROM.

Instructors provide content and learning materials in a variety of formats, which include written content materials, audio or video broadcast lectures, PowerPoint show for lecture or content delivery, and web links to content on the internet about the subject matter of the course [Abb05].

3.1.1 E-Learning Past and Present

In origin E-learning first appeared in the mid 1970s. Early Computer Based Training (CBT) consisted of mainly text-based (book-on-a-screen) type resources that were both ineffective and unpopular with learners. This led to the general concern that CBT will take the human touch away from training. Today learning edge e-learning course includes rich media including animation, video workplace scenarios, and audio, and it includes highly interactive learning activities to involve and engage the learner so it includes pre-training assessments, which increase learning effectiveness and dramatically reduce the time needed for each competency [Min03].

Today's e-learning programmers are designed to integrate with other delivery methods to create a powerful blended approach to people development [Tay02].

3.1.2 Benefits of E-Learning [Che06]

E-Learning provides benefits to individuals and organizations, the most commonly cited benefits were:

. Convenience for learners.

. The ability to reach more learners.

. The cost effectiveness of e-learning.

. The freedom e-learning provides learners to direct their own learning.

The three biggest barriers to e-learning cited by all respondents were

. Lack of funding.

. Staff time.

. Expertise.

The major benefits and barriers identified are often opposite side of the same issue: just as e-learning promises cost-effectiveness, it can be hard for organizations to fined funding to get an initial learning format promises to free up staff time once it's implemented.

3.2 Personalization

Personalization is a means of meeting the user's needs more effectively and efficiently, making interactions faster and easier and, consequently, increasing user's satisfaction and the likelihood of repeat visits[Abd06].

It provides users with the information they want or need, without expecting from them to ask for it explicitly [Gna06].

The user information can be based on history, on the actions a user takes or on information gathered through a survey [God01].

Many different research disciplines have contributed to explore personalization techniques and to evaluate their usefulness within various application areas [Bal04].

Supporting the individual user in his working, learning, or information access is one of the main goals of user modeling. Personal or group user models make it possible to represent and use information about preferences, knowledge, abilities, emotional states, and many other characteristics of a user to adapt the user experience and support [Zim05].

3.2.1 The user profile

When developing personalization applications, one of the key technical issues is the problem of how to construct accurate and comprehensive profiles of individual users that provide the most important information describing about whom those users are and how they behave [God01].

The user profile represents the information that is known about the user. In personalized systems, this information is crucial, because this is where the personalization will be based on [God01].

The objective of collecting user information is to build a profile that describes a user's interests (preference), role in an organization, entitlements, purchases, or some other set of descriptors important to implement the personalization. Classify this information in to two components [Abd06]:

1. Explicit profiling (Factual profiling): asks each user to fill out information or questionnaires by a specific form. This technique has the advantage of letting users to tell directly what they want to see.
2. Implicit profiling (Behavioral profiling): tracks the user's behavior, this technique is generally transparent to the user.

Most personalization techniques request users to input accurate personal data, and then analyze that information to make recommendations. Answering such personal questions can make people somewhat reluctant to use the personalization services [Wan02].

3.2.2 Personalization techniques to analyze collected data

When the user collected data is available, the next step is to analyze this information in order to present or recommend documents or any actions specific to the user. Making such recommendations is the most challenging step [Abd06].

Personalization techniques can be divided into three categories according to the degrees of personalization. They are rule-based filtering, Content-based filtering and collaborative filtering [Wan02].

Highly personalized services usually combine two or more of these techniques [Hel03].

A. Rule-based filtering

In a rule-based personalization engine, rules being relevant to the designed model are described and processed [Hel03].

Rule-based personalization is useful for granting the application provider direct control over the personalization logic. Fixed rules can be captured quite easily in static rule-based personalization logic. Rule-based systems however are not capable of learning using dynamic data sources to discover trends over time. It is the application provider’s task to constantly analyze rule-based systems and manually optimize and feed new rules into the system resulting in increasing complexity with a high number of rules and ultimately costly manpower investments throughout the life of the solution. Therefore within dynamic applications rule-based personalization should be combined with other techniques [Hel03].

B. Content-based filtering

It tracks user behavior and preferences, recommending items that are similar to those that users liked in the past. So it works by analyzing the content of the objects to form a representation of the visitor's interests. One example is a document filtering system that analyzes documents based on keywords. Content-based filtering is most suitable when the objects are easily analyzed by computer and the visitor's decision about object suitability is not subjective [Abd06].

The advantages of Content-Based Filtering include that this method closely pays attention to the individual user’s profile which is especially useful if the user does not fit into any simple grouping. The fact that also new items can be recommended to the user without defining new explicit rules is of particular interest to applications with large amounts of content items. Among the difficulties and limitations of Content-Based Filtering are the following: By definition, Content-Based Filtering can only suggest similar items in the same domain. With a limited range of choices recorded, this can quickly turn into an over-specialization resulting in a very narrow view [Hel03].

C. Collaborative filtering

Collaborative filtering, also called group filtering, is designed to serve relevant material to users by combining their own personal preferences with the preferences of like-minded others [Wan02].

Like the name indicates, collaborative filtering engines are able to create recommendations based on explicit and implicit "collaboration" of multiple application users. The basis for this personalization is the assumption that persons with similar tastes will also have similar interests and therefore react similarly in comparable usage situation. In practice that means that one user's profile is compared with the interest/preferences profiles of other users. Collaborative filtering in its clear form is solely based on explicit choices of other users. While, in contrast to content-based filtering, it is able to suggest items which are of a different domain, it can hardly recommend new, unpopular or very specific items due to the lack of collaborative data for these items. Also, for good results a critical amount of explicit input needs to be available to the collaborative filtering engine. In practice Collaborative Filtering and Content-Based Filtering are often combined due to their complementary qualities and weaknesses [Hel03].

3.2.3 Personalization security [Hel03]

With security we can differentiate between actual security and perceived security. Both should be oriented on the level of security required based on the weight and types of transactions and profile properties being used.

This includes ensuring security in terms of integrity (i.e. the data is not getting changed or corrupted), authentication (i.e. including a way to identify the user) and confidentiality (other people should not be able to see one user's personal information).

On the backend side, an appropriate level of data storage security is needed. Protecting the valuable User Profile data with adequate software and encryption measures should go along with the application provider ensuring a decent level of organizational and physical security. This includes redundant data storage and regular off-site backups as well as restricted virtual and physical access to the profile server limited to authorized staff only.

3.3 Personalization in distributed e-learning

Personalization is a concept with raising importance. It intends to adapt an application to the user's individual needs such that each user gets the idea that the system was created just for him/her and knows what s/he likes [God01].

The Personalization techniques are domain-independent, meaning that the techniques used are usually general and can be applied to any domain [Abd06]. E-learning is growing very fast and many universities and companies are already supporting in some way an e-learning solution [Tri03]. E-learning environments can greatly profit from the features offered by the new personalization techniques for accessing to resources, so that personalized support for learners becomes even more important, when e-learning takes place in the information networks [Dol04].

When a learning community is large and very active the amount of information and the incoming flow of new information constitute an "information overload" problem for the single learner, which becomes unable to read all the available resource and to select the resource most relevant from him/her. Personalization techniques can help the user by timely recommending specific individual information which meets his/her specific interest, searching and filtering the different resources.

E-learning can be viewed as an innovative approach for delivering well designed, learner-centered, interactive, and facilitated learning environment to anyone, anyplace, anytime by utilizing the attributes and resources of various digital technologies along with other forms of learning materials suited for distributed learning environment. Personalized learning using distributed information in learning environments is still an unsolved problem in e-learning research.

Personalization is used in e-learning to achieve a number of goals [Abd06]:

(1) Adaptation the content of e-learning resources to personal views.

(2) Ensuring of privacy.

(3) Handling of large amounts of data by lettering out what resources are of the user interests.

(4) Making the access to resources faster and easer.

Chapter Two

Distributed Database System

2.1 Introduction

A distributed database system is a collection of logically interrelated shared databases, which are physically distributed across separate nodes in a network managed by a database management system that allows the distribution to appear transparent to the end users. They share no physical components and run on sites that are completely independent of each other [Wel05].

A distributed database system is the combination of a database system, a distributed database management system and a computer network. The DBS manages data, the DDBMS is the software that manages the DB, and the computer network makes the communication possible.

The distributed database may be homogeneous or heterogeneous, in a homogeneous distributed database; all the nodes are running the same DBMS to manage the local database. In addition all the nodes use the same communications software. In contrast, in a heterogeneous distributed databases, there are differences among the nodes with respect to the running DBMS, or the communications software. [Bri03]

Users access the distributed database through local applications which do not require data from other sites or through global applications which do require data from other sites.

Figure (2.1) shows what a typical distributed system with a distributed database might look like. The client would connect to the global applications, which would in turn connect across the network to the sites on which the fragments are located, process the pieces of the application applicable to those sites and their data, and return the result set to the calling application [Wel05].

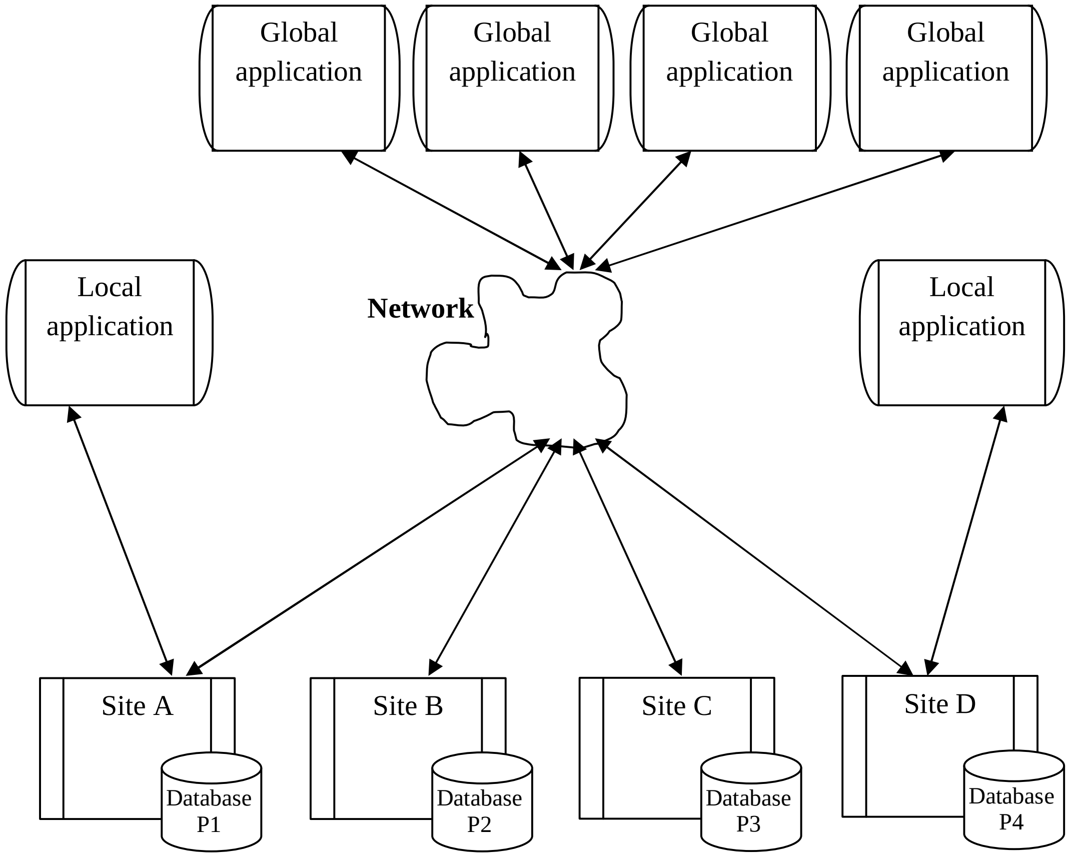
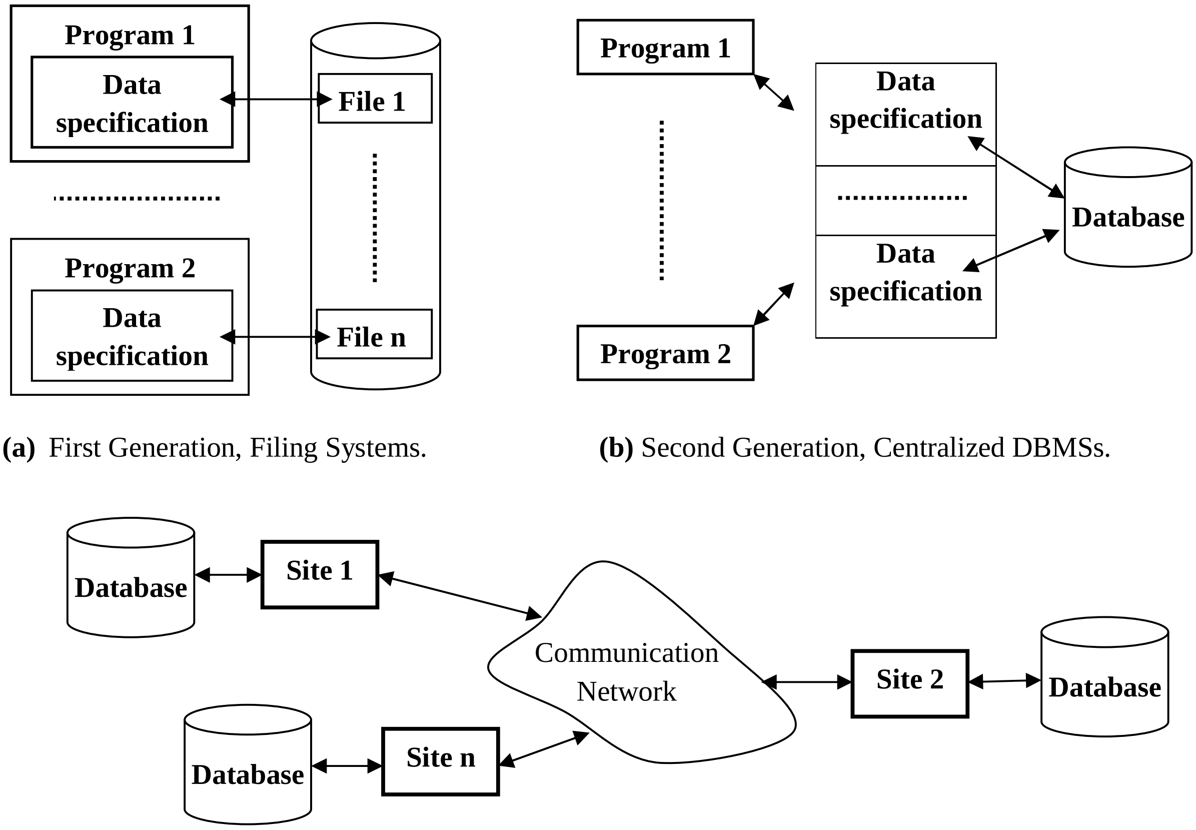


Figure (2.1) Distributed database system.

The main difference between centralized and distributed database system is that, in the former, the data reside in one single location, whereas in the latter, the data reside in several location. [Kor02]

Figure (2.2) depicts the evolution of DDBSs, the first generation of data processing is decentralized and unintegrated where data are stored in individual files and specifications are embedded into the programs that manipulate the data. Files are therefore not shared, and any changes in the file structure will affect the data specifications in the programs as shown in Figure (2.2-a). The second generation of data processing is centralized and integrated at which data are stored in a centralized database, and data specification are stored in a centralized location (normally the same location as the database) as shown in Figure (2.2-b). The advantage of this model is that changes in database may only affect data specifications but not the programs. The third generation of data processing – Figure (2.2-c) – is distributed and integrated at which data and their local specifications are distributed in a network, and there also exists a global view of all the data stored in a network [Zho01].



(a) First Generation, Filing Systems.

(b) Second Generation, Centralized DBMSs.

(c) Third Generation, Distributed DBMSs.

Figure (2.2) Evolution of DDBSs.

2.2 Database Systems

The following are some concepts related to database systems:

2.2.1 The Relational Model [Zho01]

The following term definitions related to relational database:

a. A domain D: is asset of atomic values and is associated with a type. For example, INTEGER, STRING, DATE are domains. A domain has a name, a type, and a format. For example, BIRTHDATE could be a domain's name, its type could be DATE, and its format could be dd/mm/yy.

b. A relation schema R: denoted by R(A_1, A_2, …, A_n), is made up of a relation name R and a list of attributes A_1, A_2, …, A_n. The domain of attribute A_i is denoted as dom(A_i), i=1, …, n. A relation schema is used to describe a relation. R is called the name of the relation and n the degree of relation.

c. A relation r of the relation schema R (A_1, A_2, …, A_n): denoted by r_i, is a set of m-tuples r = {t_1, t_2, …, t_m}. Each tuple t_j, j= 1, …, m is an unordered list of n values t = <v_1, v_2, …, v_n>, where each value v_i, 1<= i <= n is an element of dom(A_i), or is a special null value. m is called the cardinality of the relation.

A relation schema is an abstract kind of object, and a relation (table) is a concrete picture (instance) of such an abstract object. The primary key of a relation is a unique identifier that can uniquely identify a tuple of that relation. Table (2.1) lists some formal terms and their informal equivalents.

Table (2.1): formal and equivalent informal terms

Formal Terms	Informal Equivalents
Relation	Table
Tuple	Row or Record
Cardinality	Number of Rows
Attribute	Column or Field
Degree	Number of Column
Primary Key	Unique Identifier
Domain	Pool of Legal Values

A relational database contains relations and nothing but relations. Each of these relations has the following properties:

a. There are no duplicate tuples – so there is always a primary key.

b. Tuples are unordered.

c. Attributes are unordered.

d. All (simple) attribute values are atomic, or equivalently, relations do not contain repeating groups. A relation satisfying this condition is said to be normalized.

2.2.2 Structured Query Language [Dat90]

Structured Query Language (SQL) is used to formulate relational operation (i.e., operations that define and manipulate data in relational form). There are two types of operations:

a. Data Definition Language (DDL): including CREATE-TABLE and DELETE-TABLE.

b. Data Manipulation Language (DML): including INSERT, DELETE, UPDATE, and SELECT.

2.2.3 Relational Algebra [Dat90]

Relational algebra consists of a collection of high-level operators that operate on relations. Each of such operators takes one or two relations as its input and produces a new relation as its output. The original eight operators defined by Codd (the pioneer of the relational model) are:

a. **The traditional (mathematical) set of operations:** union, intersection, difference, and Cartesian product.

b. **The special relational operations:** restrict, project, join, and divide.

The original eight operators work as follows:

a. **RESTRICT:** Extract specified tuples from a specified relation.

b. **PROJECT:** Extract specified attributes from a specified relation.

c. **PRODUCT:** Builds a relation from two specified relations consisting of all possible combinations of tuples, one from each of the two relations.

d. **UNION:** Builds a relation consisting of all tuples appearing in either or both of two specified relations.

e. **INTERSECT:** Builds a relation consisting of all tuples appearing in both of two specified relations.

f. **DIFFERENCE:** Builds a relation consisting of all tuples appearing in the first and not the second of two specified relations.

g. **JOIN:** Builds a relation from two specified relations consisting of all possible combinations of tuples, one from each of the two relations, such that the two tuples contributing to any given combination satisfy some specified condition.

h. DIVIDE: Takes two relations, one binary and one unary, and builds a relation consisting of all values of one attribute of the binary relation that match (in the other attribute) all values in the unary relation.

2.2.4 Normalization

Normalization provides a method of representing data and their relationships precisely in a tabular format that makes the database easy to understand and operational efficient. It is basically formalization of avoiding redundancy; or trying to achieve "one fact at one place". [Dat90]

There are five normalization levels (forms). Each form's rules place additional conditions on the database design. By satisfying the first set of rules, the data is said to be in the first normal form or 1NF. Moving on to the second set results in the second normal form (2NF), and so on up to the fifth normal form (5NF). Figure (2.4) depicts these normal forms levels. The conditions addressed by each normal form are summarized as follows [Zho01]:

a. 1NF: Repeating groups.

b. 2NF: A column's dependency on only part of a composite key (partial dependency).

c. 3NF: A non-key column's representing a fact about another non-key column (transitive dependency).

d. 4NF: Two or more independent, multivalued facts occurring for an entity.

e. 5NF: Independent columns (symmetric constraints).

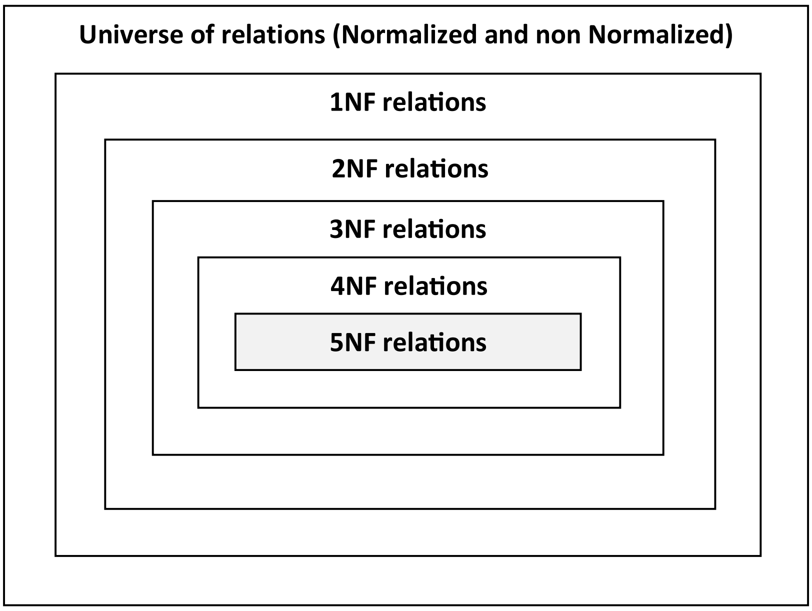


Figure (2.4) levels of Normal forms.

2.3 Advantages and Disadvantages of DDBS

Three issues lead to the need for Distributed database systems [Sey98]:

a. Distributed System (Hardware/Software). Nowadays, most of computer systems are connected by networks and therefore are distributed.

b. Distributed Application. Many applications are distributed in nature. For example, the banking system, the point of sales system, etc.

c. Distributed Data. Most data is distributed in nature. Many organizations have multiple locations. Therefore, data are already distributed.

The advantages of a distributed database system include [Wel05]:

a. It can be constructed in such a way that it can easily reflect the organizational structure.

b. Allowing for local autonomy.

c. Economics because many small computers cost less than a single big one.

d. Modular growth because the new modules are easily added to the existing system.

e. Improved sharing.

f. Improved availability so single-site failure does not make the system entirely inoperable.

g. Improved reliability.

h. Improved performance.

There are also some disadvantages in using a DDBS, some of them are listed below [Dat90, Wel05]:

a. Complexity: the distributed systems are complex at least from technical point of view. That complexity should be the implementer's problem not the users.

b. Security: is an added concern in a distributed system. This adds even more concern to the network system, in which there are increased security concerns because of its configuration and multiple points of interception.

c. Integrity: the Integrity control is more difficult in a distributed system.

d. Cost: add costs of system management and a learning curve that needs to be overcome to efficiently administer such a system.

e. Lack of standards: a considerable lack of standards in distributed database creation and administration.

f. Lack of experience: a considerable lack of experience in distributed database creation and administration.

g. The added complexity of design.

2.4 Structure of Distributed Databases [Kor91]

A distributed database system consists of a collection of sites, each of which maintains a local database system. Each site is able to process local transactions, those transactions that access data only in that single site. In addition, a site may participate in the execution of global transactions, those transactions that access data in several sites. The execution of global transactions requires communications among the sites.

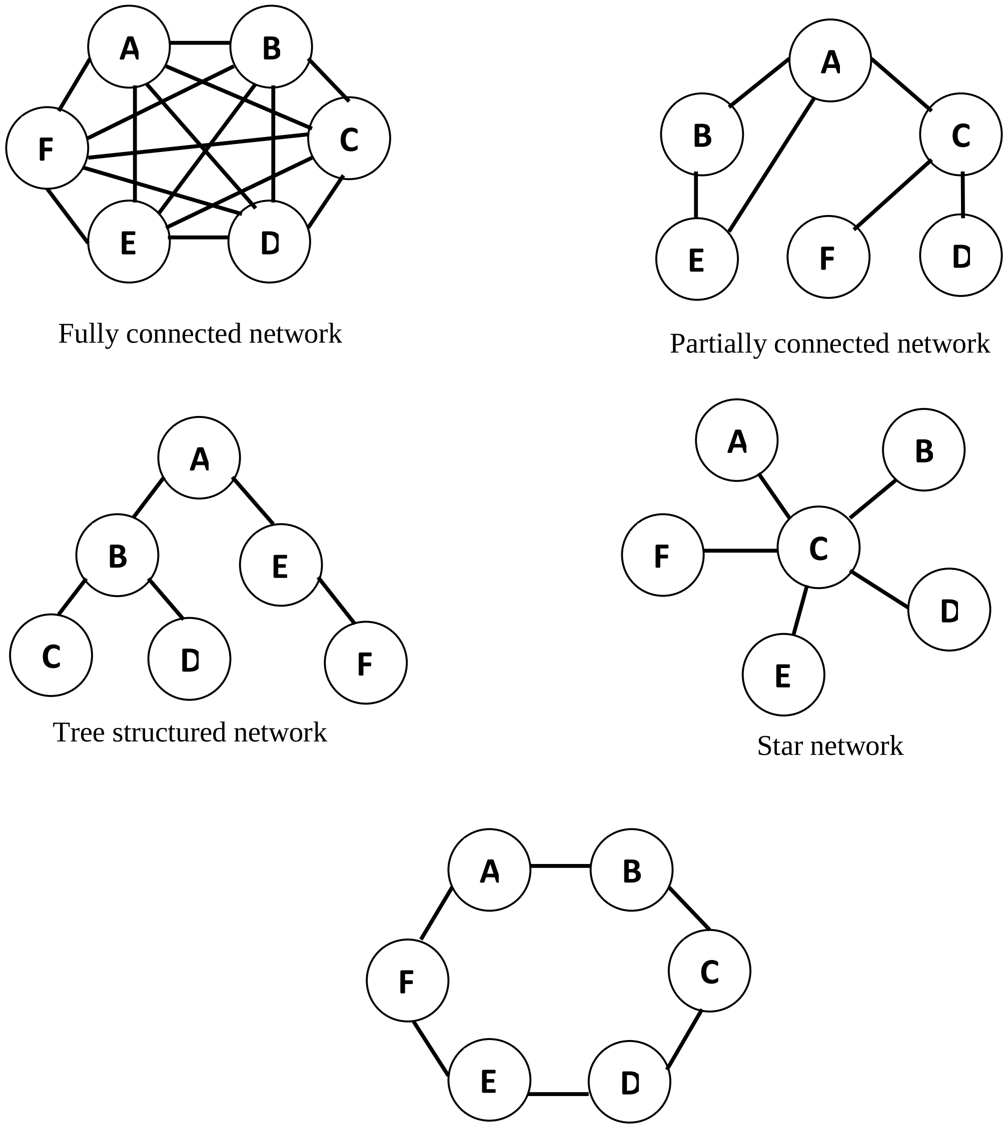
The sites in the system can be connected physically in verity of ways. The various topologies are represented as graphs whose nodes correspond to sites. An edge from node A to node B corresponds to a direct connection between the two sites. Some of the most common configurations are depicted in Figure (2.5). The major differences among these configurations involve:

a. **Installation cost:** The cost of physically linking the sites in the system.

b. Communication cost: The cost in time and money to send a message from site A to B.

c. Reliability: The frequency with which a link or site fails.

d. Availability: The degree to which data can be accessed despite the failure of some links or sites.



Fully connected network

Partially connected network

Tree structured network

Star network

Ring network

Figure (2.5) Network topology.

2.5 Distributed Database Management System

Basically, there are three dimensions for databases, DDBMSs, and applications. Figure (2.6) shows the alternatives in implementing DDBMSs [Zho01]:

a. **Distribution:** means that databases, DDBMSs, and applications can be centrally or distributed in a network.

b. **Autonomy:** means that the control of databases, DDBMSs, and applications can be done centrally, autonomous, or in between.

c. **Heterogeneity:** means that databases, DDBMSs, and applications can be of the same type or of different types.

In general, with homogenous distributed system each site runs its own copy of the same DBMS. More precisely, the DBMSs at different sites all support the same data model and the same database operations. While in a heterogeneous distributed system different DBMSs at different sites support different data models and/or different database operations. There is no point in building a heterogeneous system from scratch. But there might be advantages in connecting together a set of pre-existing centralized DBMSs to form a distributed system [Ozs00].

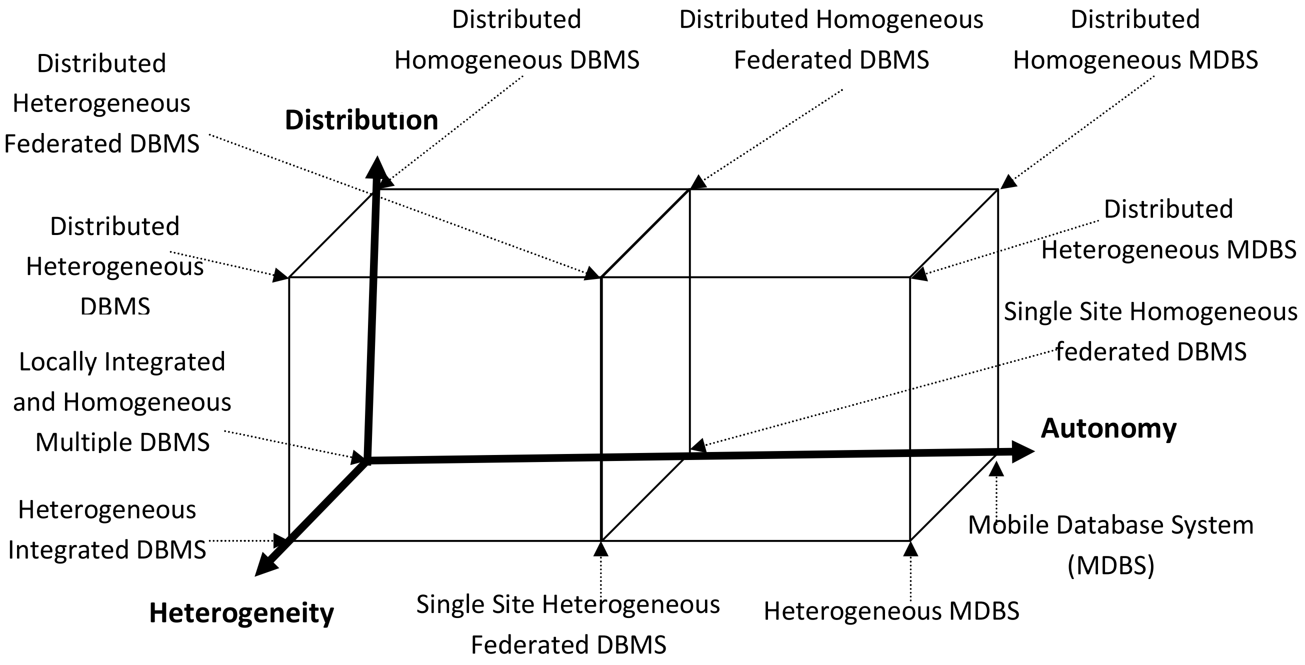

Distributed Homogeneous DBMS
Distributed Homogeneous Federated DBMS
Distributed Homogeneous MDBS
Distributed Heterogeneous Federated DBMS
Distribution
Distributed Heterogeneous DBMS
Distributed Heterogeneous MDBS
Single Site Homogeneous federated DBMS
Locally Integrated and Homogeneous Multiple DBMS
Autonomy
Heterogeneous Integrated DBMS
Mobile Database System (MDBS)
Heterogeneity
Single Site Heterogeneous Federated DBMS
Heterogeneous MDBS

Figure (2.6) DDBMS implementation alternatives.

2.5.1 DDBMS Architecture

Architecturally, a DDBS consists of a non-empty set of data sites and a (possibly empty) set of query sites. The data sites have data storage capability while the query sites do not. The latter only run the user interface (in addition to applications) in order to facilitate data access at data sites [Ozs00].

There are two points of view for the DDBMS architecture: *the level of distribution transparency* and *the alternative of client/server models*, as highlighted in the subsequent sections.

A. Level of Distribution Transparency [Cer85]

The reference architecture for a DDB is defined to represent different level of distributed transparency. Figure (2.7) shows the reference architecture for a DDB.

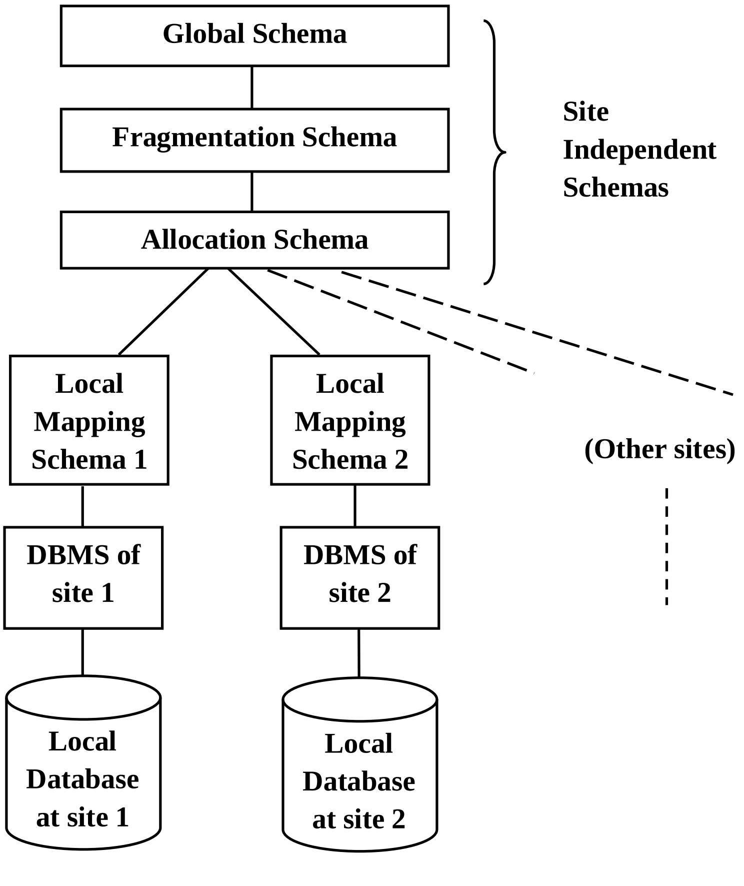


Figure (2.7) Reference architecture for distributed databases.

At the top of this architecture is the global schema. The global schema defines all the data that are contained in the DDB as if the database were not distributed at all; i.e., the global schema consists of the definition of a set of global relations.

Each global relation may be split into several non-overlapping portions that are called fragments. Fragments are logical portions of global relations that are physically located at one or several sites of the network. The mapping between global relations and fragments is defined in the fragmentation schema. This mapping is one to many; i.e., several fragments correspond to one global relation, but only one global relation corresponds to one fragment.

The allocation schema defines at which site(s) a fragment is located. The type of mapping defined in the allocation schema determines whether the DDB is redundant or non-redundant; this mapping is one to one. All the fragments that are correspond to the same global relation R and are located at the same site J constitute the physical image of global relation R at site J. there is, therefore, a one to one mapping between a physical image and a pair (global relation, site).

These three levels are site independent; therefore, they do not depend on the data model of the local DBMSs. At a lower level, it is necessary to map the physical images to the objects that are manipulated by local DBMSs. This mapping is called a local mapping schema and depends on the type of local DBMS. In a homogenous system, there is no need for the local mapping schema. While in heterogeneous system, there are different types of local mappings at different sites. This architecture provides a very general conceptual framework for understanding DDBs. The three most important objectives that motivate the features of this architecture are:

a. **Separating of concept of data fragmentation from the concept of data allocation.** This separation distinguishes two different levels of distribution transparency, namely fragmentation transparency and location transparency. Fragmentation transparency is the highest degree of transparency and consists of the fact that the user or application programmer works on global relations. Location transparency is the lower degree of transparency and requires the user or application programmer to work on fragments instead of global relation, without knowing where the fragments are located. The separation between the concept of fragmentation and allocation is very convenient in DDB design, because the determination of relevant portions of the data is thus distinguished from the problem of optimal allocation.

b. **Explicit control of redundancy.** The reference architecture provides explicit control of redundancy at the allocation level.

c. **Independence from local DBMSs,** which is called local mapping transparency.

B. The Alternative of Client/Server Models

The popular DDBMS architecture is the Client/Server Architecture, which requires full collaboration from each server and their respective client(s) [Rod05].

There are a number of different architectural models for the development of a DBMS, ranging from client/server systems (where the query sites correspond to clients and the data sites correspond to servers) to a peer-to-peer system where no distinction is made among the client machines and the server machines. These architectures differ with respect to the location where each

DBMS function is provided. In case of client/server DBMSs, the server does most of the data management work. This means that all of query processing and optimization, transaction management, and storage management are done at the server. The client, in addition to the application and the user interface, has a DBMS client module that is responsible for managing the data that is cached to the client and (sometimes) managing the transaction locks that may have been cached as well. A typical client/server functional distribution is given in Figure (2.8) [Zho01].

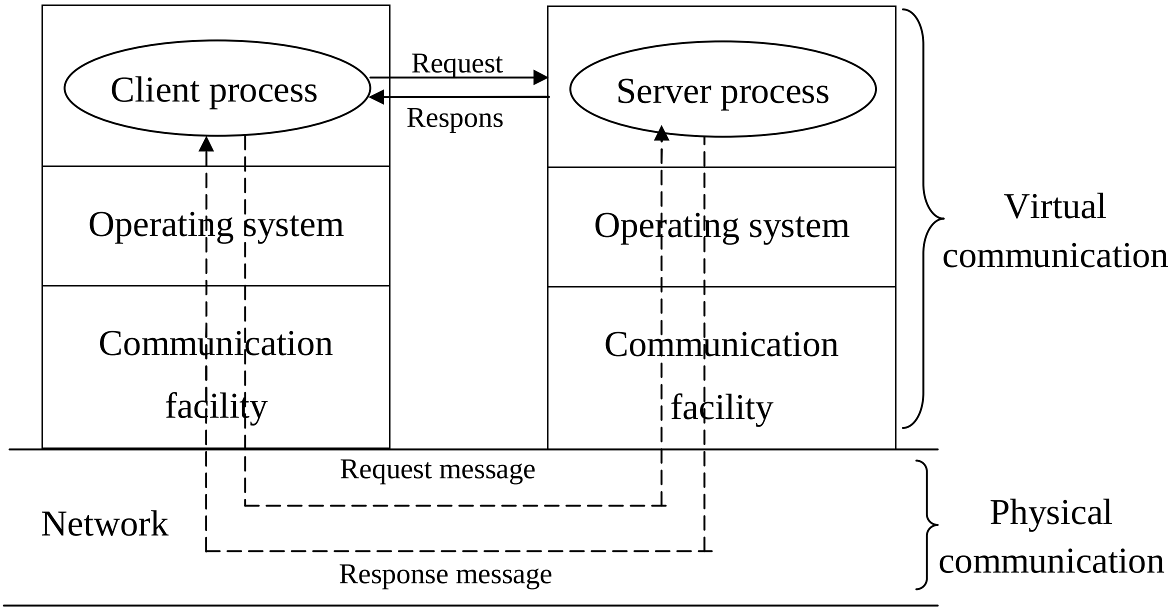


Figure (2.8) Client/Server model.

There are two types of client/server model:

a. Two-Tier client/server model: in which, the simplest architecture is a Multiple-Client/Single-Server system. From a data management perspective, this is not much different from centralized databases since database is stored on only one machine (the server), which also hosts the software to manage it. However, there are some important differences from centralized systems in the way transactions are executed. More sophisticated client/server architecture is one where there are multiple servers in the system so it called Multiple-Client/Multiple-Server model. In this case, two alternative management strategies are possible: either each DBMS client manages its own connection to the appropriate server, or each client knows only its home server (which then communicates with other servers as required). The former approach simplifies server code, but loads the client machines with additional responsibilities (heavy client), while the latter approach concentrates data management functionality at the servers and provides transparency of data access at the server interface (light client) [Ozs99].

b. Three-Tier client/server model [Zho01]: extends the basic client/server model by adding a middle tier to support the application logic and common services. In this architecture, a distributed application consists of the following three types of components:

1. User interface and presentation processing: These components are responsible for accepting inputs and presenting the results. These components belong to the client tier.
2. Computational function processing: These components are responsible for providing transparent, reliable, secure, and efficient distributed computing. They are also responsible for performing necessary processing to solve a particular application problem. These components belong to the application tier (proxy server).
3. Data access processing: These components are responsible for accessing data stored on external storage devices (such as disk drivers). These components belong to the back-end tier (server).

References

[Abd06] Abdalaal A., ***"Techniques for Personalization in E-learning***", M.Sc thesis, Department of Computer Science, Vrije University, Brussel, Belgium, 2006.

[Abb05] Abbas A., ***"Design and Implementation of Web Course Tools (WebCT)"***, M.Sc thesis, Department of Computer Sciences, University of Technology, Baghdad, Iraq, 2005.

[Bal04] Baldoni M., Henze N., and et.al, ***"Personalization Functionality for the Semantic Web: Identification and Description of Techniques"***, REWERSE, 2004.

[Bal05] Baldoni M., Baroglio C., and et.al, "***Personalization for the Semantic Web"***, Springer-Verlag, Berlin, Heidelberg, Germany, 2005.

[Bas04] Basim A., "***Recommender System for E-Commerce Data***", M.Sc. Thesis, Department of Computer Science, Al-Nahrain University, Baghdad, Iraq, 2004.

[Bri03] Briz L., ***"Implementable Models for Replicated and Fault-Tolerant Geographically Distributed Databases"***, Ph.D. Thesis, University of Politecnica, Valencia, Spain, 2003.

[Cer85] Ceri S., "***Distributed Database Principles and Systems"***, McGraw-Hill Company, 3ed printing, 1985.

[Che06] Chesemore S., "***Trends in E-Learning for Library Staff"***, OCLC Online Computer Library Center Inc., 2006.

[Dat90] Date C.J., ***"An Introduction to Database Systems"***, Addison-Wesley Publishing company, Inc., Vol. I, 5th Edition, U.S.A., 1990.

[Dol04] Dolog P., Sintek M., and et.al, "***Personalization in Distributed E-Learning Environments***", Research Center, University of Hannover, Hannover, Germany, 2004.

[Gna06] Gnasa M., ***"Congenial Web Search"***, Paper, Bonn University, Bonn, Germany, 2006.

[God01] Goderis S., "***Personalization in Object-Oriented Systems***", M.Sc Thesis, Department of Computer Science, Vrije University, Brussel, Belgium, 2001.

[Hel03] Hellmund M., ***"Smart Personalization for Wireless Applications***", Diploma thesis, Department of Digital Media, University of Applied Sciences, Furtwangen, Germany, 2003.

[Kim05] Kim J., and Lee E., ***"A Study of User XQuery Method for Personalization Recommendation in Semantic Web"***, WORLD ENFORMATIKASOCIETY, 2005.

[Kor91] Korth H. and Silberschatz A., "***Database System Concepts"***, McGraw-Hill Companies, Inc., 2ed Edition, U.S.A., 1991.

[Kor02] Korth H.F. and Silberschatz A., ***"Database System Concepts***", McGraw-Hill Companies, Inc., 4th Edition, U.S.A., 2002.

[Kos02] Kose I., ***"Distributed Database Security***", 2002.
http://www.bilmuh.gyte.edu.tr/~ispinar/BIL673/DDbsecurity.pdf.

[Min03] Ministry Training Source, ***"Frequency Asked Question (FAQ)***", 2003.
http://www.ministrytrainingsource.org/elearning.htm.

[Mit05] Mitchell C., ***"Component of DDB***", Technical Report, Department of Computer Science, Norfolk State University, 2005.

[Ozs99] Ozsu M. and Valduriez P., "***principles of distributed database systems***", Prentic Hall, 2ed Edition, 1999.

[Ozs00] Ozsu M.T.," ***Distributed Database"***, 2000.
http://WWW.pmg.lcs.mit.edu/~chmoh/pubs/DDB.pdf.

[Reu06] Reus B., ***"Customization and personalization in the Internet Economy***", M.Sc. Thesis, University of Amsterdam, Amsterdam, Holland, 2006.

[Rod05] Rodriguez E., ***"WAMDAS: A Web Service-Based Wireless Alarm Monitoring and Data Acquisition System for Pharmaceutical Plants"***, M.Sc. Thesis, Department of Computer Engineering, University of Puerto Rico, 2005.

[Sey98] Seydim A., "***An Overview Of Distributed Database Management***", 1998. http://www.pmg.lcs.mit.edu/papers/DDBM.pdf.

[Tay02] Taylor R., ***"Introduction to E-learning"***, 2002. http://www.creativelearningmedia.com.

[Tri03] Trifonova A., ***"Hoarding content in M-learning context"***, Ph.D Thesis, Department of Information and Communication Technologies, University of Trento, Italy, 2003.

[Wan02] Wang J., and Lin J., ***"Are Personalization Systems Really Personal? – Effects of Conformity in Reducing Information Overload***", Department Of Information Management, National Chung-Cheng University, Taiwan, 2002.

[Wel05] Wells A., ***"Grid database design"***, Taylor & Francis Group, U.S.A., 2005.

[Zho01] Zhou W., "***Distributed Database Systems"***, 2001. http://WWW.cs.adfa.oy.au/teaching/studinfo/dsd/DistributedDBS.pdf.

[Zim05] Zimmermann A., Specht M., and et.al, "***Personalization and Context Management***", Kluwer Academic Publishers, Printed in the Netherlands, 2005.