A FRAMEWORK FOR A PERSONALIZED E-LEARNING SYSTEM

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

EBRU ÖZPOLAT

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY
IN
ELECTRICAL AND ELECTRONICS ENGINEERING

JUNE 2009

Approval of the thesis:

A FRAMEWORK FOR A PERSONALIZED E-LEARNING SYSTEM

submitted by **EBRU ÖZPOLAT** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy in Electrical and Electronics Engineering Department, Middle East Technical University** by,

Prof. Dr. Canan Özgen ____________________
Dean, Graduate School of **Natural and Applied Sciences**

Prof. Dr. İsmet Erkmen ____________________
Head of Department, **Electrical and Electronics Engineering**

Prof. Dr. Gözde Bozdağı Akar ____________________
Supervisor, **Electrical and Electronics Engineering Dept., METU**

Examining Committee Members

Prof. Dr. Semih Bilgen ____________________
Electrical and Electronics Engineering Dept., METU

Prof. Dr. Gözde Bozdağı Akar ____________________
Electrical and Electronics Engineering Dept., METU

Prof. Dr. Adnan Yazıcı ____________________
Computer Engineering Dept., METU

Prof. Dr. Petek Aşkar ____________________
Computer Education and Instructional Technologies Dept., Hacettepe University

Assist.Prof.Dr. İlkay Ulusoy ____________________
Electrical and Electronics Engineering Dept., METU

Date: June 09, 2009

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name : Ebru ÖZPOLAT

Signature :

ABSTRACT

A FRAMEWORK FOR A PERSONALIZED E-LEARNING SYSTEM

Özpolat, Ebru

Ph.D., Department of Electrical Electronics Engineering
Supervisor: Prof. Dr. Gözde Bozdağı Akar

June 2009, 150 pages

This thesis focuses on three of the main components of an e-learning system: Infrastructure model, data integration and personalization. For the infrastructure model, our aim is to get best use of heterogeneously structured, geographically distributed data resources. Therefore, a detailed analysis of the available infrastructure models is carried out and an open source reference implementation based on grid technology is implemented. Furthermore, a simple data integration mechanism is proposed for the suggested reference implementation.

For personalization, a statistical algorithm is proposed based on extracting and utilizing the learner model. The learner model based on Felder-Silverman learning style is extracted automatically using NBTree classification algorithm in conjunction with Binary relevance classifier for basic science learners. Experimental results show that the performance of the proposed automated learner modelling approach is consistent with the results, obtained by the questionnaires traditionally used for learning style assessment.

In the thesis, the classification results are further utilized for providing the user with personalized queries.

Keywords: Interactive learning environments; personalization in e-learning

ÖZ

KİŞİSELLEŞTİRİLMİŞ ELEKTRONİK ÖĞRENME SİSTEMLERİ İÇİN BİR ALTYAPI

Özpolat, Ebru

Doktora, Elektrik Elektronik Mühendisliği Bölümü

Tez Yöneticisi: Prof. Dr. Gözde Bozdağı Akar

Haziran 2009, 150 sayfa

Bu tez, elektronik öğrenim sistemlerinin temel parçalarından üçüne odaklanmıştır: Mimari model, veri birleştirme ve kişiselleştirme. Mimari model için amacımız, coğrafi olarak dağıtık ve farklı yapıda oluşturulmuş veri kaynaklarından en fazla şekilde faydalanabilmektir. Bu amaçla, mevcut mimari modeller detaylıca incelendi ve grid teknolojisine dayalı, açık kaynak kodlu, referans mimari yapıları yeniden oluşturuldu. Buna ek olarak, önerdiğimiz referans mimari yapı için veri birleştirme mekanizması önerdik.

Elektronik öğrenme sistemlerindeki kişiselleştirme konusu için kullanıcı modeli oluşturmaya ve kullanmaya dayalı statik bir algoritma önerdik. Bu tezde Felder-Silverman öğrenme stilini temel alan bir kullanıcı modelini otomatik olarak oluşturuyoruz. Bu kullanıcı modeli temel bilimler için çalışan kulanıcıları hedef alır. Bu kullanıcı mdeli, NBTree algoritmasını Binary Relvance sınıflandırıcısıyla bilikte kullanarak oluşturulur. Deneysel sonuçlar önerilen kullanıcı modeli metodunun performansının öğrenme stili teşhisinde kullaılan klasik anket sonuçları ile uyumlu olduğunu göstermektedir.

Bu tezde, ayrıca, kullanıcıya kişiselleştirilmiş sorgular sağlamak için sınıflama sonuçlarından yararlanılmıştır.

Anahtar Kelimeler: Etkileşimli öğrenme ortamları, elektronik öğrenmede kişiselleştirme

To My Mother

ACKNOWLEDGMENTS

I would like to thank my supervisor Prof.Dr. Gözde Bozdağı Akar for her supervision, encouragements and support throughout the development of this thesis.

Also, I am grateful to my examining committee members for their suggestions and reviewing of this thesis.

I would like to thank ASELSAN Inc. for supporting academic studies.

Finally, I wish to thank my parents for their morale support and giving me the strength and courage to complete this thesis study and my brother for his valuable suggestions and support throughout this thesis.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ANN	Artificial Neural Network
BR	Binary Relevance
CORDRA	Content Object Repository Discovery and Registration/Resolution Architecture
DRI	Digital Repository Interoperability
DT	Decision Tree
FSLSM	Felder Silverman Learning Style Model
GDQS	Grid Distributed Query Service
GUI	Graphical User Interface
GLOB	Grid Learning Object
HBDI	Herrmann Brain Dominance Instrument
ICT	Informaion and Communications Technology
ID3	Iterative Dichotomiser 3
ILS	Index of Learning Styles
IMS-LD	Instructional Management Systems Learning Design
kNN	k Nearest Neighbour
LMS	Learning Management System
LP	Label Powerset
LSAL	Learning Services Architecture
LSDOs	Learner Selected Data Objects
LSI	Learning Style Inventory
MBTI	Myers-Briggs Type Indicator
MLE	Managed Learning Environment
NB	Naive Bayes
OASIS	Open Architecture and Schools in Society
OGSA	Open Grid Service Architecture
OGSA-DAI	Open Grid Service Architecture Data
OGSA-DQP	Open Grid Service Architecture-Distributed Query Processor
OKI	Open Knowledge Initiative

OQL	Object Query Language
QES	Query Evaluation Service
PT	Problem Transformation
RIO	Reusable Information Object
SIF	Schools Interoperability Framework
SOA	Service Oriented Architecture
SRW	Search/Retrieve Web Service
SQI	Simple Query Inetrface
SVM	Support Vector Machine
WSRF	Web Services Resource Framework
XML	Extended Markup Language

CHAPTER 1

INTRODUCTION

1.1. Motivation of the Thesis

Today's technology in information systems changes also the state of the art in education. There is a tremendous trend towards e-learning systems to access the huge, available and various information worldwide [1]. There are certain key parts in e-learning system: Infrastructure model, data integration, personalization.

E-learning applications vary in scale and purpose. In general, e-learning applications are delivered via learning management systems (LMS) or 'virtual learning environments' that are structured around course delivery to short, just in time targeted, single purpose learning experiences, delivered in context. In LMSs, context is derived from the LMS's knowledgebase and applicable only within the LMS. However, recently, the question is that, does e-learning require any purpose built application software like LMS or is it an activity that is defined by engagement with information and communications technology (ICT) more broadly. Furthermore, what distinguishes e-learning activities, if at all, from other activities associated with learning, education, or training – such as scholarly research, information seeking, information browsing, sense-making, or knowledge sharing – especially when many of these activities are enabled by information and communications technology (ICT) infrastructure? [4]. Additionally, when context is provided by the infrastructure, then contextualization information can be shared by all tools. Therefore, e-learning is now facilitated by an increasing range of specialised e-learning applications within the wider infrastructure [4]. Further, there is significant investment in many e-learning communities in developing infrastructure models supporting e-learning available.

Some of the examples, explained in detail in the following chapter, are namely Learning Services Architecture, Content Object Repository Discovery and Registration/Resolution Architecture(CORDRA), Digital Training Systems Architecture, Open Architecture and Schools in Society (OASIS), Open Grid Services Architecture(OGSA), Open Knowledge Initiative (OKI) etc. [4].

With this motivation, although there are various open source or commercial LMS platforms; like Moodle, Dokeos, Blackboard Inc. or eCollege; in this thesis work, we concentrated on e-learning systems, which are not constructed via LMSs. Moreover, in this thesis work, we concentrated on how to access geographically distributed e-learning resources. Hence, an infrastructure model is required for accessing and searching heterogeneously structured and geographically distributed information. An infrastructure model can be described generally as a framework, combining software and protocols in such a way that provides both the flow and processing of information. Such an infrastructure model should be used with an appropriate query interface to process user's queries.

Regarding an infrastructure model, there are many commercially available e-learning platforms and systems based on client-server (DIESEL), peer-to-peer (e.g. Edutella) or Web services (e.g OLAT) architectures. However, these systems have some limitations in scalability, storage capability and distribution of computing power. To support e-learning systems in scalable, open, dynamic, heterogeneous environments, there is a trend to use grid technology in e-learning systems [21], [22], [23].

Another key issue in e-learning is query interfacing, which covers how the user pose queries and retrieves the results. Regarding the query interface part, there are some of the available approaches in the literature are as presented in [8],[9],[10],[11] and explained in detail in the following chapter.

Furthermore, another main component of an e-learning system is providing data integration capability. When the information resources are heterogeneously structured and geographically distributed, data integration capability is required to improve interoperability among resources by establishing a relationship between data resources. Data integration can be provided using various approaches depending on the application, like global schema usage, schema mapping or schema matching, peer-to-peer data integration, ontology-based data integration. Therefore, query interfacing should be designed such that, it should cover data integration of resources and reflect tis in the results of submitted queries.

Once the infrastructure model is established and heterogeneous resources are integrated, the user can pose queries and excess the information as he/she wishes. However, the user should be able to benefit from all of these resources in a best efficient way, such that whenever they pose a query to get the necessary information regarding their research, they should get the most appropriate results for them. This requirement address the personalization capability issue; i.e. supporting individual differences and needs of learners, in an e-learning system.

The available e-learning systems and online search engines used for educational purposes typically do not care of personalization and treat all learners in the same way regardless of their personal needs and characterization. Actually, each learner has individual needs and differences like prior knowledge, learning style, cognitive style, competencies, motivation, communication preferences, and so on. Therefore, personalization in an e-learning system plays a dominant role by providing enhancement in learner's learning experience in terms of relevancy, motivation, effectiveness, efficiency, and satisfaction. The personalization issue can be achieved by capturing and utilizing the learner model. Learner models can be extracted based on personality factors like learning styles, behavioral factors like user's browsing history and knowledge factors like user's prior knowledge.

Additionally, recently, educational studies have concentrated on personality factors like learning styles, their effect on learning process and how they can be contributed in technology enhanced learning process. The personality factor, learning style arises from the educational and psychological theories, which claims that learners have different ways, in which they prefer to learn. According to Felder [30], [73], learners with a strong preference for a specific learning style may have difficulties in learning, if teaching style does not match the preferred learning style. Hence, it can be concluded that, including learning styles of learners into e-learning systems, enhances learner's learning efficiency in comparison to e-learning systems without supporting learner's learning style.

Further, traditionally, learning style of a learner can be diagnosed using well investigated instruments like long questionnaires, whose reliability and validity is verified [39]. However, the drawback of using questionnaires or other tests in diagnosing learning styles is that if the tests are too long or the learners are either unaware about the future usage of the questionnaire results or in a bad, unconcentrated mode, then learners tend to choose answers arbitrarily without thinking carefully about them. Therefore, the obtained results can be inaccurate or may not reflect the actual learning style. As a result, new trend in this area is to diagnose learning style of the learner automatic learner modeling.

There are various methods in the literature for automatic learner modeling [44]-[70]. These methods differ based on the attributes they use (personality factors, behavioral factors etc.), the underlying technique (Bayesian networks, decision trees etc.) and the underlying infrastructure (LMSs, special user interface etc.). Most of these methods are developed specific to an underlying LMS using the participation of the learner in forums, chats and mail systems etc. In addition, there is a lack in designing an automatic learner model to diagnose learner's learning style using only the content of the data objects selected by the learner but not learner's other behavior observed in time scale. Such kind of an automatic learner model design will bring the independency from a background, underlying LMS. As a result, simplicity and easy integrability of such a learner model to a LMS, just Web based search or any infrastructure model can be obtained.

1.2. Contributions of the Thesis

Based on the motivation issues presented in the previous section, the contributions of this thesis can be stated as follows:

- A detailed analysis of infrastructure models is carried out and based on it; the implementation of the e-learning system based on OGSA is established.
- In order to get best use of geographically distributed and heterogeneously structured data resources, query interfacing needed for data integration is proposed. The proposed data integration module, used in query interfacing, overcomes language heterogeneity between heterogeneous, geographically distributed databases using a simple mapping algorithm for data integration.
- To provide personalization in e-learning systems, a novel automatic learner modelling technique is proposed. The proposed approach diagnose learning styles based on Felder-Silverman learning style model using NBTree classification in conjunction with a multi-label classification method, namely Binary Relevance (BR) classifier. The success of the proposed novel approach is presented with the experimental results. The proposed approach contributes to fulfill the lack in the literature in two ways. First, the approach to diagnose learning styles uses only the content of the data objects selected by the learner but not learner's other behavior observed in time scale. Secondly, the proposed approach is able to handle common data objects selections of learners with different learning styles. In the proposed approach, initially the learner interest is collected explicitly using generic queries. Then, the learner profile is constructed using a conversion unit based on keyword-mapping. The learner model is built by processing the learner profile over a clustering unit and then using a decision unit. The proposed automatic learner model is portable such that it can be integrated to a LMS, a Web based search engine or an infrastructure model.

1.3. Structure of the Thesis

The thesis is organized in 8 chapters.

In the next chapter, firstly, the key features of an infrastructure model are introduced and then various infrastructure models and suitable query interfaces available in the literature are investigated. After that grid-based OGSA infrastructure model is mentioned. Then a detailed literature review is given on grid applications for e-learning. Finally, the implementation of the e-learning system based on OGSA is presented in two parts. First, we introduced the required components and their role in the infrastructure model. Secondly, we explained the realization of the OGSA infrastructure model and OGSA-DQP GUI as query interface in detail. The big picture of the proposed, complete e-learning system based on OGSA is presented at the end of this chapter.

In chapter 3, another component of an e-learning system, namely data integration capability for query interfacing is covered. Under this scope, first a detailed literature review of grid applications and data integration is presented. Second, we explained our contribution regarding data integration capability for query interfacing of the proposed e-learning system, namely the proposed data integration module. Finally, we conclude this section with an application scenario over the proposed data integration module.

In chapter 4, we concentrated on personalization based on learner modelling, another main component of e-learning systems. It is addressed by the literature review regarding the works related to learner modelling in e-learning and then a detailed analysis on learning styles, the basis of the proposed learner model, is presented.

In chapter 5, since we used classification in the proposed automatic learner modeling approach, we presented the investigation of the most important classification algorithms for text data in the literature and different methods for multi-label classification problem with evaluation metrics.

In chapter 6, we described the proposed automatic learner model with implementation details. Furthermore, experimental results and the criticism of experimental results are presented.

In chapter 7, a sample application scenario for the overall system and also application of the proposed automatic learner model to web is illustrated.

Chapter 8 concludes the thesis by highlighting the contributions and discussing limitations and future directions.

CHAPTER 2

INFRASTRUCTURE MODEL

In this chapter, first we present the key features of an infratructure model and give a detailed analysis and comparison of the available infrastructure models and query interfaces in the literature. After analysis, we concentrated more on grid environments, since grid services extend web services by stateful services, service instantiation, named service instances, lifetime management and a base set of capabilities including discovery facilities [107]. We present an implementation of an open source infrastructure model combined with a query interface.

2.1. Infrastructure Model

In an infrastructure model, the following key features should be available [4].
- ***standards dependency:*** An infrastructure model should be a clear model supported by a community to satisfy some purposes including;

- defining a clear overview of the e-learning space as it applies to a community;
- defining a map for implementation of standards and specifications;
- supporting strategic decisions relating to infrastructure and standardization;
- documenting an agreed implementation set of technologies;
- providing an implementation and development environment, etc.

- ***monolithic application silos/ service oriented architectures (SOAs):*** An infrastructure model can be based on either monolithic application silos or SOAs. SOAs provide delivery by Web services and are trendy in infrastructure models due to its major benefits like lower cost, reusability, and rapid development.

An infrastructure model can be developed on a traditional distributed database or on Grid, where the Grid community adopted an open architecture based on the concepts of soa's and Web Services.

- ***networks and connectivity:*** Networks and connectivity play a crucial role in the development of a successful infrastructure model. Accordingly, connectivity should be not taken just an enabler of networks and access to content.Connectivity is also an organizing principle in its own right, since connections enable relationships (involving people and services) and relationships enable networks.From the technical perspective, an infrastructure model can be based on a peer-to-peer architecture or on an advanced broadband network (client-server architecture) or wrapper-mediator architecture.

There are various infrastructure models available in the literature. Some of them are as follows [4]:

Carnegie-Mellon LSAL: Learning Services Architecture : The focus of the learning services architecture is based on a flexible design supported by open standards covering learning technology standards and common web and network standards. These standards are used for information exchange, behavior descriptions and component integration. Hence, it provides interoperability of components and learning content. In other words, both the **Learning Services Architecture** and the **Learning Services Stack** together provide a framework for developing service-based and component-based learning technology systems.

Carnegie-Mellon LSAL: Content Object Repository Discovery and Registration/Resolution Architecture – CORDRA : CORDRA is designed to be an enabling model to bridge the worlds of learning content management and delivery, and content repositories and digital libraries. CORDRA aims to identify and specify (not develop) appropriate technologies and existing interoperability standards that can be combined into a *reference* model.

IBM LOTUS eLearning Infrastructure: This infrastructure is an application software that offers two ways to deliver training via the Web: either in real-time, virtual classrooms or in self-paced online modules.

IMS Global Consortium – Abstract Framework: This framework does not define the IMS architecture, rather it is an abstract representation of the set of services that are used to construct an eLearning system in its broadest sense. It is a framework that covers the possible range of eLearning architectures that could be constructed from the set of defined services.

JISC e-learning Framework: The JISC has already made a considerable contribution to the development of the Managed Learning Environment (MLE) concept and supporting interoperability standards. However there is still a lack of pedagogical flexibility and innovation in the design of e-learning tools, environments and architectures to support institutional processes such as effective e-learning. The JISC e-Learning Programme aims to solve this issue and concentrates on four main themes: e-learning and pedagogy; technical framework and tools for e-learning; innovation and distributed e-learning.

Open Architecture and Schools in Society (OASIS): This project aims to promote small virtual communities in schools within the public educational system with the main role of socialization.

Open Grid Services Architecture (OGSA) : The aim of the OGSA Working Group is to achieve an integrated approach to future OGSA service development via the documentation of requirements, functionality, priorities, and interrelationships for OGSA services.

Open Knowledge Initiative (OKI): It is focused to provide new learning technology in higher education via collaboration among leading universities and specification and standards organizations. The result of this collaboration is an open and extensible architecture that specifies how the components of an educational software environment communicate with each other and with other enterprise systems.In other words, OKI provides a modular development platform for building both traditional and new applications while leveraging existing and future infrastructure technologies.

SIF – Schools Interoperability Framework: The Schools Interoperability Framework (SIF) is a non-profit membership organization and aims to create a set of platform independent, vendor neutral rules and definitions, called the *SIF Implementation Specification*, to enable software programs from different companies to share information.

Considering the architecture and editability (open source code/not open source code), we derived the following comparison table as presented in Table 2.1.

Table 2.1 Comparison of Infrastructure Models

Infrastructure Model	**Source code**	**Architecture**
Carnegie-Mellon LSAL: Learning Services Architecture	Not open source	web services
Carnegie-Mellon LSAL: Content Object Repository Discovery and Registration/Resolution Architecture – CORDRA	Not open source	A reference model, web services
IBM LOTUS eLearning Infrastructure	Not open source	web services
IMS Global Consortium – Abstract Framework	Not open source	web services
JISC e-learning Framework	Not open source	web services
Open Architecture and Schools in Society (OASIS)	Open source	web services
Open Grid Services Architecture (OGSA)	Open source	grid services
Open Knowledge Initiative (OKI)	Open source	web services
SIF – Schools Interoperability Framework	Not open source	SIF-Implementation Specification-web services

As presented on Table 2.1, only the infrastructure models, namely, OASIS, OGSA and OKI are open source. Further, among them, only OGSA is based on grid services, which is our other selection criterion. Since grid services extend web services by stateful services, service instantiation, named service instances, lifetime management and a base set of capabilities including discovery facilities [107]. In section 2.3, we will also give other e-learning systems for grid environment. The details of OGSA will be given in the following subsection.

2.1.1. OGSA

OGSA provides a scalable, dynamic, heterogeneous environment via grid technology. Considering,

- OGSA based applications use as middleware, a de facto standard, open source Globus Toolkit. So, it provides multi-purpose protocols and interfaces that address fundamental issues like authentication, authorization, resource discovery and access, which are currently not supported by Internet and Web infrastructures.
- OGSA enables remote sources to be used for two purposes, namely for data retrieval tasks and for computational tasks.
- OGSA provides systematic access to heterogeneous remote data and computational resources.
- Moreover, as its name implies OGSA is based on SOAs and can integrate and coordinate distributed resources and users without a centralized control and provides mechanisms to monitor network connections.

we propose to use OGSA [5], [6] in this thesis.

On the other hand, considering alternatives to OGSA, another work [7] concludes that there could be alternative software stacks for OGSA-based Grids. The motivation of this work is that, although OGSA has been a major step forward for Grid Computing, its de facto reliance on the Web Services Resource Framework (WSRF) and WSNotification have left some in the its developer community questioning if the Grid software and protocols have become too complex. Therefore, some people make a feasibility work for an alternative software stack, based on WS-Transfer/WS Eventing. According to qualitatively and quantitatively evaluations of each approach namely WSRF/WS Notification and WS-Transfer/WS Eventing, the result is that these two approaches are overwhelmingly equivalent in their functionality and implied performance. There are slight differences like the naming and the creation of resources, which does not affect our infrastructure model selection.

The next step is to provide the communication of the learner with the proposed infrastructure model for his queries. This aim is achieved via query interface. The detailed analysis of query interfaces is presented in the following section.

2.2. Query Interface

There are various approaches in the literature for providing query interface [8]-[11].These different approaches are presented below:

- ***Simple Query Interface (SQI) :*** SQI is an API for querying heterogeneous repositories of learning resource metadata. It is developed as a part of CEN/ISSS Workshop on Learning Technology efforts to improve interoperability between learning resource repositories. Considering its structure, SQI is composed of a set of methods used by a source to connect a target system, to configure and submit queries to this target system and retrieve results from this target system.

  Furthermore, SQI specification is based on work by Ariadne, CELEBRATE, Edutella, Elena, EduSource, ProLearn, Universal/EducaNext and Zing. Taken in turn, that means a venerable network of repositories and e-learning tools, a network brokerage system that ties the VLEs of schools together, a peer-to-peer educational resource tool network, an agent based 'smart space' project, the Canadian network of learning object repositories, a new network of excellence on professional learning, a learning object brokerage system, and the next generation web services based specs derived from the Z39.50 library protocol.

- ***IMS Digital Repository Interoperability (DRI) specification:*** The DRI specification provides five core commands (search/expose, gather/expose, alert/expose, submit/store, request/deliver) on a highly abstract level. As a result, the implementers have to make too many design choices to manage the interoperability problem. Regarding the query languages, the specification recommends both Z39.50 (with its own query language) and XQUERY query languages.
- ***Z39.50-International: Next Generation (ZING):*** ZING developed a protocol called SRW (Search/Retrieve Web Service). SRW has many similarities with SQI, but also some differences. Regarding the differences, SRW is a web-based protocol and transfers information in XML format over http. Moreover, SRW takes advantage of "Common Query Language" (CQL), a powerful query language, which is a human –readable-string-query-representation. However, the main drawback in SRW is that it only supports synchronous queries. That means, a source system sends a query to a target, and waits until the target returns a response ; i.e. the results, if at all possible.
- ***OGSA-DQP GUI:*** It can be used on Grid environment with OGSA reference implementation namely OGSA-DQP(Open Grid Service Architecture-Distributed Query Processor). OGSA-DQP GUI provides the user the ability to import OGSA-DQP resources and launch queries over distributed heterogeneous structured data sources wrapped by OGSA-DAI using OQL (Object Query Language). It provides interaction with OGSA-DQP services as an alternative to the command-line client. OGSA-DQP GUI can be considered as a corresponding API of SQI in OGSA based systems using OGSA-DQP in a restricted way such that this GUI provides the submission of requests and presenting the corresponding results.

Comparing these approaches, it can be concluded that SQI is independent of query languages and metadata schemas. Therefore, there appears the requirement of the design for a wrapper and wrapping service interface at the SQI interface. Moreover, in addition to SQI API and wrapper mechanism, to complete interoperability between learning resource repositories, there is a need for a common query schema and results format as a part of the semantic model of an educational network in systems benefiting from SQI API. Regarding DRI specification, there is a lot of work to be completed by the designer between the data source layer and query interface commands. On the other hand, ZING has similarities with SQI regarding its usage requirements; i.e. required work to be fulfilled by the designer. However, considering OGSA-DQP GUI, on Grid environments, it is used with OGSA-DQP. Hence, the designer does not need to work for a wrapping mechanism and accessing data resource schemas, which are already managed by OGSA-DQP.

2.3. Review of e-Learning Systems Based On Grid Technology

As mentioned in section 2.1, there are e-learning systems based on grid technology [21], [22], [23]. In this section, we present the detailed analysis of how these e-learning systems get use of grid technology.

One of the major works in this area is [21]. In this work, an architecture based on grid technology for e-learning is outlined and the structure of a grid learning object is introduced as a first step for the realization of this novel approach. The proposed architecture for e-learning grid is composed of two parts. One is the core grid middleware and the other is Learning Management System (LMS). An LMS coordinates all learning-related activities; like storing and managing learning objects, providing discussion boards or forums about a teaching subject and so on.

In the proposed architecture, LMS is integrated into core grid middleware via additional grid services and operation calls. The LMS interacts with the grid middleware in such a transparent way that a learner is not aware of the grid. The only required thing for a learner is a Java-enabled Internet browser to use both the LMS and the grid [21].

Further, the interface between LMS and grid is obtained via GLOB. GLOB combines both the functionalities of traditional learning object and of grid. The proposed notion of global learning object (GLOB) to be used in e-learning grid is as presented on Figure 2.1.

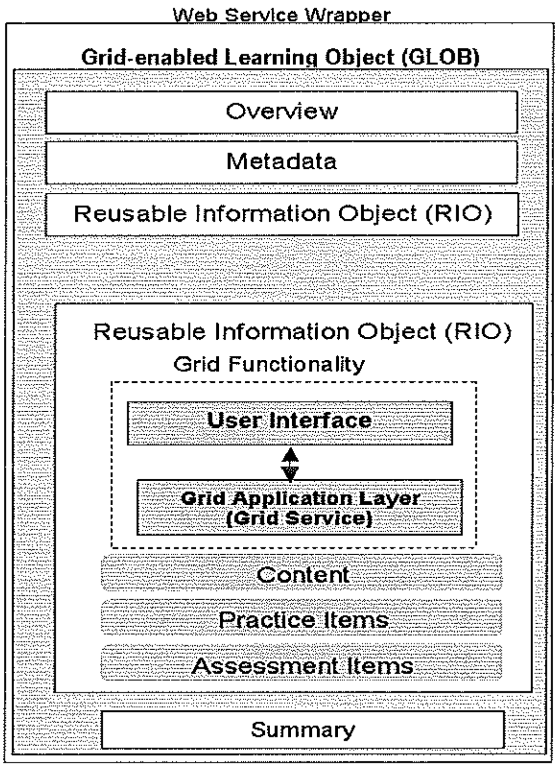


Figure 2.1 Structure of a Grid Learning Object (GLOB) as in [21]

The GLOB has the following parts: An overview of the lesson, metadata used to find GLOB,several RIOs and summary.A RIO contains both contents in different parts, practice items and assessment items and grid functionality implemented as grid service and accessed via a user interface.

Another work [22], proposes an e-learning platform using grid environment, built up using a number of open source e-learning softwares in a heterogeneous environment. The proposed architecture is presented in Figure 2.2.

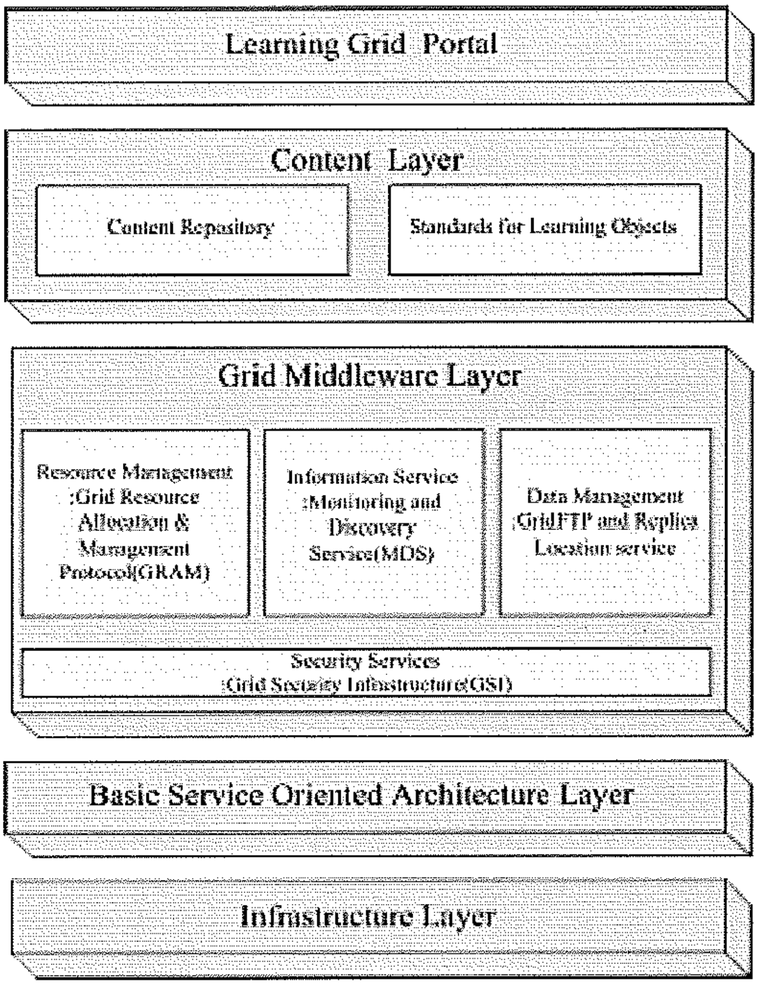


Figure 2.2 An E-Learning Grid Architecture given in [22]

The infrastructure layer, at the bottom, is responsible for basic networking environment including hardware and software. On top of this layer comes the basic service oriented architecture layer, where implementation of basic web services related protocols like XML, UDDI/SOAP/WSDL etc. is realized. Also, elementary connectivity,reliability and flexibility for the layers on top of it is provided at this layer. The grid middleware layer handles the basic grid functionalities like resource management, information service, data managemet and security issues. The content layer on top of grid middleware layer is resonsible to store all of learning content of the e-learning system. Finally, the learning grid portal stands for supporting sign on the system.

This work describes the execution flow of a learning platform in Grid as explained below:

1) Grid portal has a user database where the user information and access rights are hold. Whenever a learner tries to login the grid portal, username and password are checked from this user database.
2) After a successful login, the list of available resources in the grid with their types and status information is presented to the learner. Also,if each computer has its own LMS, then the status information is required.
3) A broker is assigned to handle distribution of computation and data regarding requests among other resources available in the grid.
4) GridFTP is used to access the other computer's resource for the data distribution.
5) Replica Location Service supports multiple locations for the same file throughout the grid.

Another work [23], proposes some alternatives to combine grid technology and teaching learning processes based on IMS Learning Design (IMS-LD). The IMS-LD specification is a powerful resource for designing collaborative and reusable complex learning units in different environments unlike other specifications that concentrate only in packaging contents [23]. Furthermore, GLOB of the previously introduced work [21] is taken as reference and the design of GLOB is modified to achieve communication with the "IMS Learning Design Engine".

Regarding the evaluation of these works, the common point between all of these works is that combining grid with e-learning provides the possibility to incorporate big volumes of information as well as to use tools that require great computing power. However, all of these works should be further studied to improve them; like either providing the incorporation of grid as special services in [23] or in general, providing atomicity over grid or recovery protocols that help restore an operational state after a grid failure.

2.4. Implementation of The E-learning System Based On OGSA

For the infrastructure, as stated before, we rebuild the reference implementations, namely OGSA-DAI-WSRF2.1 [76], [77] and OGSA-DQP3.0 [78]-[80] according to our configuration settings. For query interface part, we rebuild OGSA-DQP GUI [11] according to our configuration settings. First, we introduce the required components with their role for the implementation. In the second part, we explain how to realize this infrastructure model according to our configuration settings.

2.4.1. Required Components For The Implementation of OGSA

For the implementation, we need the following components: Jakarta-Tomcat Web Server, Globus Toolkit, OGSA-DAI WSRF2.1, OGSA-DQP 3.0 and OGSA-DQP GUI.

Jakarta-Tomcat Web Server

The Jakarta Project offers a diverse set of open source Java solutions and is a part of The Apache Software Foundation (ASF) which encourages a collaborative, consensus-based development process under an open software license.

Jakarta is organized into subprojects and one of them is Tomcat. Tomcat is the servlet container (called CATALINA) of our infrastructure model.

Globus Toolkit

The Globus Toolkit is the implementation of the main Grid middleware. The Globus Toolkit is a community-based, open-architecture, open-source set of services and software libraries that support Grids and Grid applications. The toolkit addresses issues of security, information discovery, resource management, data management, communication, fault detection, and portability. It is packaged as a set of components that can be used either independently or together to develop applications [74].

OGSA-DAI WSRF2.1

OGSA-DAI is a middleware product. It provides access to data resources like relational or XML databases via web services. Moreover, using an OGSA-DAI web service one can query, update, transform and deliver data. In addition, OGSA-DAI web services can be used to provide web services that offer data integration services to clients.

OGSA-DAI web services can be deployed within a Grid environment. OGSA-DAI thereby provides a means for users to Grid-enable their data resources.

OGSA-DAI provides web services compliant with two popular web services specifications, namely Web Services Inter-operability (WS-I) and Web Services Resource Framework (WSRF).

Regarding the motivating issues for emergence of OGSA-DAI the following requirements can be presented:

- Grid-enable different types of data resources - including relational, XML and files
- Enable querying, updating, transforming and delivering data via web services.
- Access to data in different type of data resources in a homogeneous way.
- Access metadata about data, and the data resources in which this data is stored.
- Support the integration of data from various data resources.
- Provide web services that can be combined into higher-level web services that can be used in data integration and distributed query processing issues.
- Improve application-specific data analysis and processing issues, which are adopted by scientists for future works instead of some technical issues like handling data location, data structure, data transfer and integration.

We can state, what the middleware product OGSA-DAI supports as follows:

- Different types of data resources - including relational, XML and files - can be exposed via web services. A number of popular data resource products are supported.
- Data within each of these types of resource can be queried and updated.
- Data can be transformed (using XSLT), compressed and decompressed (using ZIP and GZIP compression).
- Data can be delivered to clients, other OGSA-DAI web services, URLs, FTP servers, GridFTP servers, or files.
- Requests to OGSA-DAI web services have a uniform format independent of the data resource exposed by the service. (Though the actions specified within each request may be data resource-specific).
- Information on the data resources exposed by an OGSA-DAI web service and the functionality supported by the service can be accessed by clients.
- OGSA-DAI users can extend OGSA-DAI web services to expose their own data resources and to support application-specific functionality, in addition to that provided by OGSA-DAI.

OGSA-DQP 3.0

OGSA-DQP is a service-based distributed query processor, which combines data access with analysis, since it is able to execute queries in parallel over OGSA-DAI data services and other (Web) services on the Grid. Regarding this ability, *OGSA*-DQP adopts techniques from parallel databases to provide implicit parallelism for complex data-intensive requests and utilizes distributed computational resources to execute queries. Further, as a service-based distributed query processor [81],

- OGSA-DQP abstracts the user from the requirement of an in-depth knowledge of the Grid technologies. This results also in an increase in the variety of people who can form requests over the Grid.
- OGSA-DQP decreases development times for the Grid programming tasks that can be expressed as database queries.
- OGSA-DQP provides efficient task execution more likely, since it combines OGSA properties with DQPs optimization and parallelism.

- OGSA-DQP employs GDS (Grid Data Services) generic interface to access remote databases instead manually constructed wrappers,
- OGSA-DQP has dynamic structure such that it is not strongly coupled to any storage system,
- In OGSA-DQP, metadata are kept in main memory and constructed on a per session basis, which provides more accurate metadata retrieval.
- In OGSA-DQP, communication between service instances occurs in the form of XML documents transmitted over SOAP/HTTP.
- OGSA-DQP is an open source tool .Hence it is open to application specific improvements.

The working principle of OGSA-DQP is completely service oriented as mentioned above. It presents its functionality as a service in the system. Furthermore, it executes queries by using a set of services like evaluation services, OGSA-DAI services and Web services.

The two services that are the fundamental components comprising the service-based DQP framework are presented as follows:

- **Grid Distributed Query Service (Coordinator).** The Grid Distributed Query Service (GDQS), or coordinator, is the main interaction point for the clients. GDQS is responsible for the processes, namely compiling, optimizing, partitioning and scheduling distributed query execution plans over multiple execution nodes in the Grid. Therefore, in order to realize these processes, when a coordinator is instantiated and set up, it is able to get the metadata and available computational resource information. The implementation of the coordinator is based on a previous work, namely on the Polar* distributed query processor for the Grid by encapsulating its compilation and optimization functionality. Now, the implementation of the coordinator can be described as a set of OGSA-DAI data service resources and activities.
- **Query Evaluation Service (Evaluator).** The Query Evaluation Service (QES), or evaluator, is coordinated by the coordinator. Each QES's

responsibility is to evaluate a partition of the query execution plan assigned to it by the coordinator. A set of evaluators participating in a query form a tree through which the data flows from leaf evaluators which interact with Grid data services, up the tree to reach its destination.

Since the coordinator is itself implemented as an OGSA-DAI data service, it can be discovered and invoked in the same way as other OGSA-DAI data services. As a consequence, OGSA-DQP provides uniform interfaces and interaction semantics for declarative request formulation that complement existing approaches to service orchestration.

Figure 2.3 provides an overview of the interactions for a whole OGSA-DQP work, covering the instantiation and set-up of an OGSA-DQP coordinator, the procedure for receiving a query and processing it via a set of evaluators. In the figure, the 3-dot sequence can be read as `and so on, up to'.

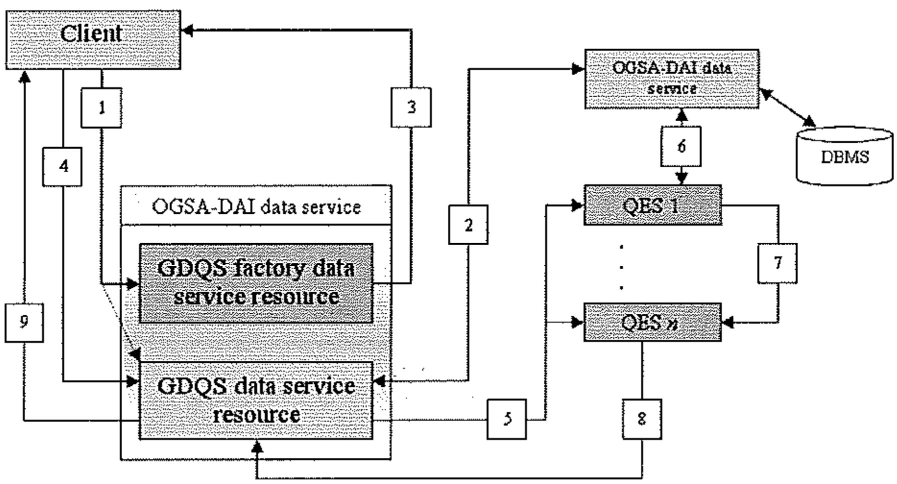


Figure 2.3 Setting up and executing queries using OGSA-DQP

1: An OGSA-DQP coordinator, a GDQS consists of two types of OGSA-DAI data service resources: GDQS factory data service resources and GDQS data service resources. Initially, when an installed coordinator is instantiated, only a deployed GDQS factory data service resource will appear. Then, using this factory data service resource a configured data service resource is created via a perform document. The created GDQS data service resource will be the interaction point to clients to execute queries.

From the technical point of view, the creation process of a GDQS data service resource is as follows: The client interacts with the GDQS factory data service resource by sending OGSA-DAI perform document which specifies that a `DQPFactory` activity should be executed. The `DQPFactory` activity is able to interact with a GDQS factory data service resource in order to dynamically deploy a GDQS data service resource. The `DQPFactory` activity is parameterized by an XML document which specifies exactly how the deployed GDQS data service resource should be configured. Configuration parameters include the databases and evaluators which can be utilized by the data service resource which is to be created. The result of this interaction is that a GDQS data service resource is created and initialized. The coordinator service now exposes this dynamically deployed GDQS data service resource and it is automatically assigned a resource ID by OGSA-DAI.

2: As a part of the initialization of the GDQS data service resource, the schemas of the databases it will use are imported by contacting the OGSA-DAI data services which wrap these databases.

3: The client receives the resource ID assigned to the created GDQS data service resource by OGSA-DAI in the result of the perform document submitted in step 1. The client will use this resource ID to identify the GDQS data service resource in subsequent interactions with this data service resource. In addition, at this point, the client is able to obtain the description of the imported database schemas.

[**Note**] Steps 1-3 represent an initialization and set-up process required to configure a GDQS data service resource. If there exists already a GDQS data service resource, which imports the databases and analysis services required by a client, then skip these steps. Each GDQS data service resource has the ability to process multiple concurrent queries and the GDQS data service resource is not terminated by a client following a query session.

4: The client submits a perform document containing a query. Queries are written in OQL and are executed by the `OQLQueryStatement` activity. The GDQS data service resource uses the Polar* query compiler to parse, optimize and schedule the query. A query plan is created, consisting of a number of partitions. Each partition specifies an individual evaluator's role in the query plan.

5: Query partitions are sent to the relevant evaluator services.

6: Some evaluators interact directly with OGSA-DAI data service to obtain data.

7: Other evaluators may interact with other evaluators to implement their role in the execution of the query.

8 - 9: Results propagate back from the evaluators to the coordinator and eventually back to the client.

<u>OGSA-DQP GUI</u>

OGSA-DQP GUI , as the application layer of the system provides the user the ability to import OGSA-DQP resources and launch queries over distributed heterogeneous structured data sources wrapped by OGSA-DAI using OQL (Object Query Language). It provides interaction with OGSA-DQP services as an alternative to the command-line client.

2.4.2. Realization Of The Infrastructure Model

Our system is composed of three machines with Windows XP operating system. The coordinator, GDQS is deployed on one of the computers. Hence, this computer is the main interaction point of the system with the client. Evaluators, QES's are deployed on all of the three computers. In this way, we provide that all of the computers can take part in evaluation of submitted queries. Moreover, we have to configue the coordinator in order to determine which evaluators and data sources to use in the system. How to configure our system and the configuration constraints are explained in details in the Appendix-A.

Regarding the distributed data sources part, on all of the machines, MySQL database management systems are installed. On these database management systems, the sample databases and their tables are created. These will be our data sources for query processing. The selection of the DBMS is restricted to OGSA-DQP supported databases. OGSA-DQP supports formally only relational databases using OQL query language.

The installation/deployment procedure till the application layer is done on all three computers as described in details in Appendix-A. However, as an overlook, the order of the deployment procedure can be summarized as follows. Considering the deployment process onto Tomcat, first of all we deploy the Globus Toolkit. The Globus Toolkit's core services, interfaces and protocols allow users to access remote resources as if they were located within their own machine room. Meanwhile, the control regarding who can use resources at what time is provided via this toolkit. In the system, it is used as a middleware for developing Grid applications. After that, we deploy OGSA-DAI WSRF 2.1 web services to Tomcat. OGSA-DAI, as a middleware product provides a homogeneous access to data resources, taking part in the system.

Next, OGSA-DQP 3.0 is deployed onto Tomcat. OGSA-DQP manages distributed query processing.

As the application layer, the special version of OGSA-DQP GUI; that involves our proposed data integration module and automatic learner model module can be used.

The general overview of the whole proposed e-learning system with these components is shown in Figure 2.4.

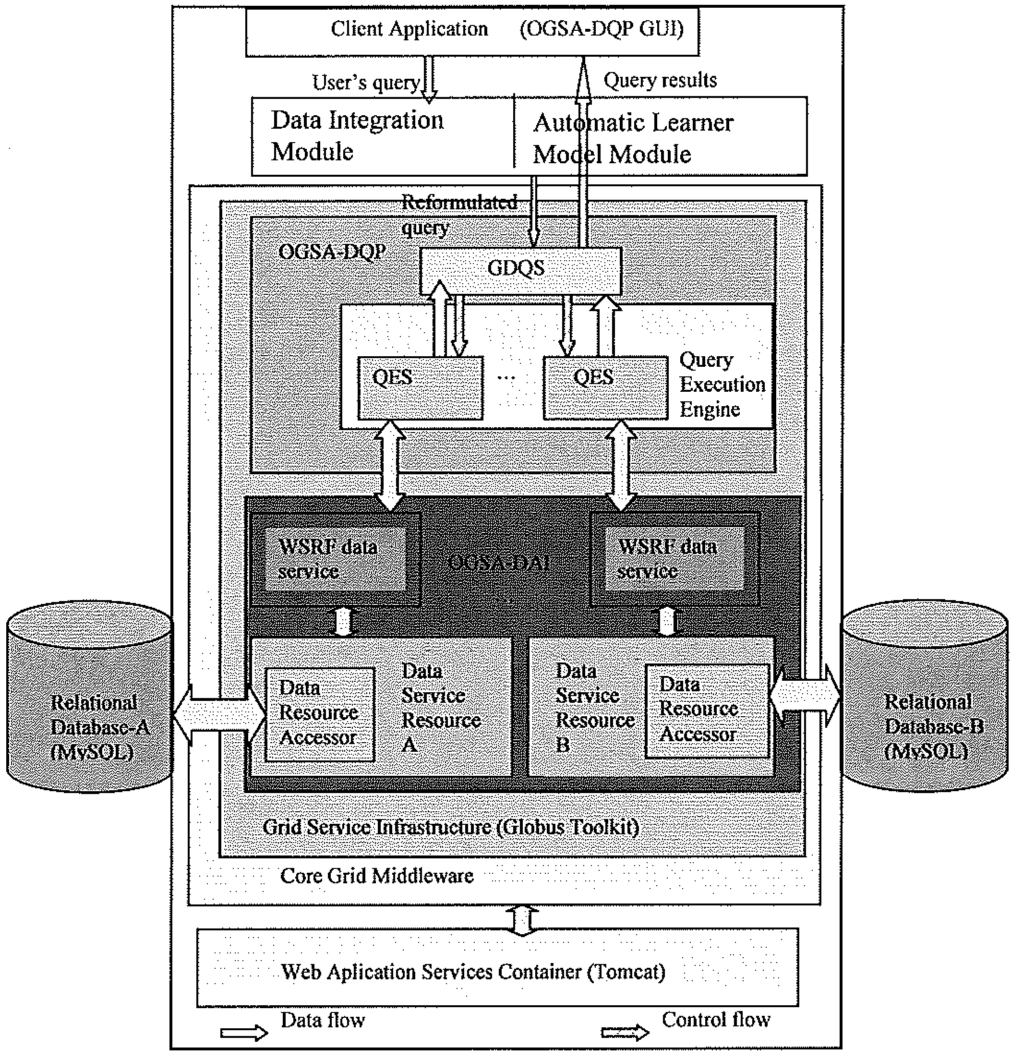


Figure 2.4 Proposed E-learning System Overview

2.5. Comparison of E-learning Systems Based On Grid Technology

Considering the e-learning systems based on grid technology, we can summarize that, [21] integrates a LMS into core grid middleware. [22] proposes an e-learning platform in grid environment using a number of open source e-learning softwares in a heterogeneous environment. [23] uses grid technology, but it is dependent on IMS-LD specification.

Regarding the evaluation of these works, the common point between all of these works is that combining grid with e-learning provides the possibility to incorporate big volumes of information as well as to use tools that require great computing power. However, all of these works should be further studied to improve them; like either providing the incorporation of grid as special services in [23] or in general, providing atomicity over grid or recovery protocols that help restore an operational state after a grid failure.

On the other hand, the proposed work utilizes the advantages of the previously mentioned works. Additionally, the proposed work uses OGSA, which includes by itself grid middleware. However, in [22], grid middleware is constructed separately. Furthermore, other works use custom-built application software like LMS or open source e-learning softwares. The proposed work is free of purpose built application software.

CHAPTER 3

DATA INTEGRATION

In this chapter, first, we present related works regarding grid applications and data integration available in the literature. Then, we introduce our contribution for data integration issue, namely data integration module for the implemented infrastructure model.

3.1. Grid Applications and Data Integration

Data integration is required to improve interoperability between heterogeneously structured data resources by establishing a relation ship between data resources. In this part, first we give a summary of the available related work regarding data integration both based on Grid and non-Grid environments to gain viewpoints about general approaches in this area. After that, we investigate some of the related works based on OGSA infrastructure model more closely. Finally, we conclude with the investigation of the related work regarding e-learning in grid environment.

One of the most general approaches in data integration systems is using global schema [12]. A global schema is like a bridge between the user and data sources by providing a uniform mediation layer. In data integration systems; global schema; i.e. virtual representation of the whole data stored at sources, is the main interaction point with the user and the heart of the system by accepting user query requests and retrieving the related results from the sources. In these systems, the relationship between the source schema, which represent data at source, and the global schema is established using mappings. Furthermore, interaction in such systems is done using a common query language using wrappers; i.e. software modules. To improve this kind of systems, integrity constraints can be expressed over the global schema.

Another approach in data integration systems is using schema mapping based on source-to-target mapping rules. These mapping rules can be generated semi-automatically like in Clio Project [13] or can be based on predefined source or target schemas.

Moreover, schema matching approaches can be used for data integration. There is also a system called COMA [14], which combines flexible schema matching approaches. In this work, simple matchers and hybrid matchers are implemented and tested. Moreover, the effectiveness of using single matchers and different combination of matchers are evaluated. Regarding simple matchers, string matchers, namely affix, n-gram, edit distance and soundex and also User Feedback, synonym and Data Type matchers are implemented. The supported hybrid matchers are as follows: Name and Type Name, which are element-level matchers and structural matchers; i.e. Name Path, Children and Leaves.

Furthermore, regarding semantic heterogeneity, data integration systems can be based on ontologies, where ontology can be described as explicit, formal descriptions of concepts and their relationships that exist in a certain universe of discourse, together with a shared vocabulary to refer to these concepts.

Another approach is peer-to-peer data integration. Data integration is decentralized and achieved without using any hierarchical structure for establishing mappings among the autonomous peers.

There are also some available data integration works on Grid environments based on OGSA infrastructure model, like AUTOMED [15]-[17], GDIS [18], [19] and GDMS [20].

The AutoMed (Automatic Generation of Mediator Tools for Heterogeneous Data Integration) Project at Birkbeck and Imperial Colleges, London, is focused on developing a toolkit to provide the transformation and integration of data from heterogeneous data sources like structured (e.g. relational), semi-structure (e.g. XML and RDF) and text data sources. Regarding this purpose, they define transformation pathways between each data source schema and an Integrated Schema [15].

The GDIS framework is focused on the development of a decentralized network of semantically related schemas that enables the formulation of queries over heterogeneous and geographically distributed data sources. For this purpose, they build data integration utilities as Grid Services to both enable users to specify semantic mappings (in the form of XMAP documents) among a set of data sources and to execute the XMAP query rewriting algorithm. In this way, it becomes possible combining or transforming data from multiple heterogeneous sources to obtain integrated or derived views of data [18], [19].

GDMS (Grid Data Mediation Service) is a data integration system, that provides a global schema namely Virtual Data Source (VDS) based on the wrapper-mediator approach [20].

When comparing data integration works over OGSA compliant environments, it is obvious that using global schema as in GDMS and integrated schema as in AutoMed brings some drawbacks with. Due to its central architecture, the design and maintenance of global/integrated schemas is difficult. It brings the limitation to local data resources not to change frequently or seriously, since such a change may break up the defined mapping rules between the local schema and the global schema. As a result, a bottleneck will occur in the system.

Regarding the work GDIS, where local mappings are used, it is good that the system benefit from the dynamic nature of the Grid systems. However, it is hard to maintain the coordination of local mapping changes occurring simultaneously. Moreover, the performance time should be larger than a system, where instead of local mappings at each site, general schema mappings are done at the beginning of the query. Furthermore, this system supports only XML data sources.

3.2. Proposed Data Integration Module

This component is prepared by us for data integration to overcome language heterogeneity between heterogeneous, geographically distributed data resources using the implemented architecture. Hence, we improve interoperability between heterogeneously structured, distributed data resources and provide query reformulation to get more comprehensive, related query results. This module is embedded to the OGSA-DQP GUI open-source software.

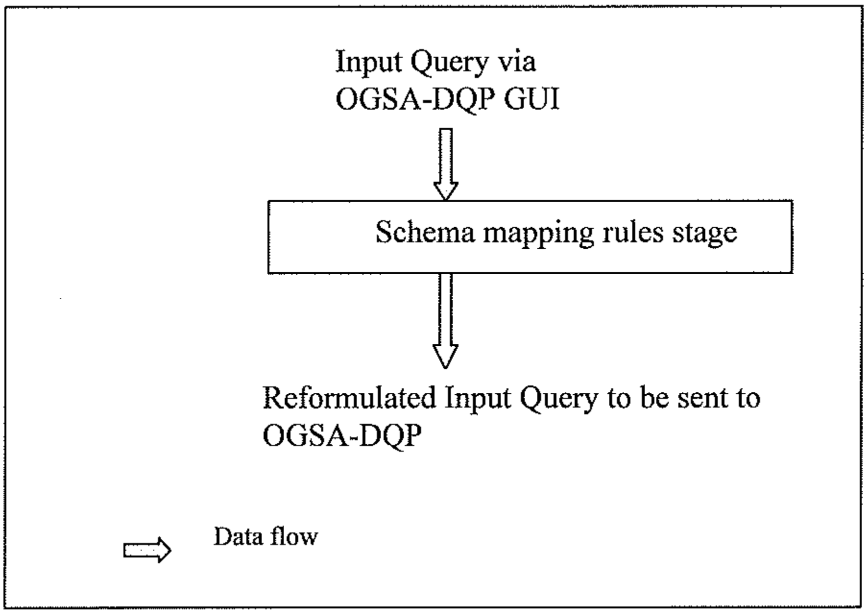


Figure 3.1 Data Integration Module

Data Integration module is shown on the Figure 3.1. The query is reformulated based on schema mapping rules. The motivating issue regarding this state is as follows: The reference implementations used in our system together; i.e. OGSA-DAI and OGSA-DQP provide us a union of local schemas. However, still there is not any relationship between these data resources and as a result the interoperability and consequently, data integration between these resources is not established. On the other hand, we want to benefit the dynamic, systematically addressing and distributed nature of Grid systems. With this motivation, instead of changing the mechanisms for data access and importing local schemas abilities of these reference implementations, we used schema matching between the imported local schemas to be queried to provide data integration over distributed data resources, which are hold with different local schemas.

The schema mapping rules stage is composed of four parts:

1) our XML document for schema mapping
2) our XML-DOM parser
3) our input query parser
4) our query reformulator including mapping process

First of all, an XML document is used to hold the schema mapping information of the imported local schemas. Then, we required an XML-DOM parser, to get the required schema information whenever necessary. Therefore, we developed our generic methods, namely parseXMLFile, readElements, readAttribute, to retrieve the required mapping information from this XML document.

On the other hand, we have to parse the input query to find out which parts to map. For this purpose, a query parser is prepared. Its generic methods, namely querySFWDivider, selectParser, fromParser, whereParser, are used to generate the input (source) for mapping process from the input query submitted by the user.

Finally, with the input for mapping process and mapping information in hand, the mapped (target) query parts are obtained according to our mapping process using our query reformulator methods, namely subdivideQuery, buildTargetQueryParts, getReformulatedSelectClause, getReformulatedFromClause, getReformulatedWhereClause. After that, the original and mapped query parts of the input query are combined in an efficient way into a new query expression. This will be our reformulated query.

As a result, the reformulated query is send to the system to be executed instead of the original query. When the results are received, the user will get all related results for his/her query. Accordingly, the procedure realized in this module can be summarized as follows:

1. Get input query
2. Parse input query for source database and table names and column names.
3. Using XML document find candidate mapped databases/tables/column names.
4. Select the mapped databases/tables/column names according to the requirements stated in the input query.
5. Construct the reformulated query in Object Query Language (OQL) by combining both the original query parts and the mapped query parts.
6. Send the reformulated query to the OGSA-DQP system.

The template of the required XML document for schema mapping:

<MAPPING>
 <source_table_name source_table_name="destination_table_name"
  source_column_name="destination_column_name></ source_table_name >
</MAPPING>

In this XML document,

- the expression *source_column_name="destination_column_name"* appears for each column name of source schema which has a corresponding column name on the destination schema.
- If a source table maps to more than one destination tables, then for each destination table a new mapping line is inserted to this XML document.

3.3. Application Scenario Over The Proposed Data Integration Module

The workflow inside the proposed module is explained in detail with the following application scenario.

Assuming that we have the following database tables as our local data resources;

```
CREATE TABLE A
(
          TITLE VARCHAR(200) NOT NULL default '',
          CATALOG_ID VARCHAR(200) NOT NULL default '',
          KEYWORDS VARCHAR(200) NOT NULL default '',
          LEARNING_PLACE VARCHAR(50) NOT NULL default '',
          LEARNING_RESOURCE_TYPE VARCHAR(500) NOT NULL
          default '',
          PRIMARY KEY(CATALOG_ID)
) TYPE=MyISAM;
```

```
CREATE TABLE B
(
          BASLIK VARCHAR(200) NOT NULL default '',
          KATALOG_NO VARCHAR(200) NOT NULL default '',
          ANAHTAR_KELIMELER VARCHAR(200) NOT NULL default '',
          OGRENIM_YERI VARCHAR(50) NOT NULL default '',
          BILGI_KAYNAK_TIPI VARCHAR(500) NOT NULL default '',
          PRIMARY KEY(KATALOG_NO)
) TYPE=MyISAM;
```

The XML document for schema mapping appears as follows:

```
<MAPPING>
 <A_A=”B” TITLE="BASLIK" CATALOG_ID="KATALOG_NO"
KEYWORDS="ANAHTAR_KELIMELER" LEARNING_PLACE="OGRENIM_YERI"
LEARNING_RESOURCE_TYPE="BILGI_KAYNAK_TIPI"></A >
</MAPPING>
```

1) Assume an input query in OQL syntax as follows:

   “%print select p.TITLE from p in A where p.KEYWORDS=”Nucleus”;”

2) Using tokens in query parser, we obtain

   “select clause” = p.TITLE

"from clause"= p in A

"where clause" = p.KEYWORDS="Nucleus"

Then, using selectParser method of query parser and subdivideQuery method of query reformulator, we obtain from "select clause", source table pointer as "p" and source_select_column_name as "TITLE"

Similarly, using fromParser method of query parser and subdivideQuery method of query reformulator, we obtain from "from clause", the source table name as "A" corresponding to table pointer "p" found from "select clause".

Using whereParser method of query parser and subdivideQuery method of query reformulator, we obtain from "where clause" the source_where_column_name "KEYWORDS" and store corresponding expression value ("Nucleus") and relation (=).

3) Using XML – DOM parser, find candidate mapped databases/tables/column names from the document obtained by parsing our XML document.
4) Build mapped (target) query parts; i.e. the mapped table name, mapped _select_ column _name and mapped_where_column_name; using the source table name, source_select_column_name and source_where_column_name obtained at step 2. For this example;

   source table name = A

   mapped table name = B and mapped table pointer = m

   source_select_column_name = TITLE

   mapped _select_ column _name = BASLIK

   source_where_column_name KEYWORDS

   mapped_where_column_name = ANAHTAR_KELIMELER

5) Combine the original query and mapped query in a single query expression by eliminating the duplicates.

   Reformulated query expression:

   "%print select p.TITLE, m.BASLIK from p in A, m in B where
   p.KEYWORDS="Nucleus" and
   p.KEYWORDS=m.ANAHTAR_KELIMELER;"

6) Send the reformulated query to the OGSA-DQP.

CHAPTER 4

PERSONALIZATION

One of the main requirements from an e-learning system is providing personalization capability. Hence, in this chapter, first, we presented the literature review regarding the works related to personalization especially based on learner modelling in e-learning. It is followed by the background information and literatur review about learning styles, which will be the basis of our proposed learner modelling for personalization. Then, we investigated classification algorithms and multi-label classification techniques to decide which one to use in our proposed automatic learner model design. Finally, we presented our contribution to personalization for the proposed e-learning system, namely automatic learner model module.

4.1. Personalization in E-Learning

Personalization, i.e. to support learner's diversity and individual needs; like prior knowledge, learning style, cognitive style, competencies, motivation, communication preferences, and so on, is one of the most important features of an efficient e-learning system. Several studies exist in literature, which address personalization in an e-learning system and elsewhere [44]-[70]. Most of these studies achieve personalization using learner/user modelling. Learner model can be described as a combination of personality factors, behavioural factors and knowledge factors [59],[68]. Furthermore, as specified in [66], in the literature, in learner modelling probabilistic and statistical methods are used to represent the personalized context like user's behaviour, like his goals or preferences. However, there are several challenges to these statistical and probabilistic methods for user modelling. [66] summarizes these challenges mainly into three groups,namely limitations of current user modeling approaches, dynamic nature of user modeling data, and efficiency considerations. In the first challenge; i.e. limitations of current user modeling approaches; this work focused on collaborative and content-based approaches used in statistical user modeling. The main limitation of collobarative approaches is that, individual user's behaviour is not predicted from specific observations of his beaviours.Instead behaviours of a mixed group of people are used to predict individual user's behaviour.On the other hand, in content-based approaches, the selection of features used in content-based model limits the usefulness of a user model. More explicitly, specific features works well for repetitive behaviours and generic features have uncertain characteristics about its usefulness in behaviour prediction. Further, the other challenge arises, when considering, that the user's behaviour changes over time and there may be new users and new subjects to introduce. All of these factors comprise the dynamic nature of user modelling. This dynamic nature becomes a problem in sparse data problem and conceptual drift techniques,used in statistical modelling. The sparse data problem, related to cold start problem, addresses the cases, where there is not enough data or observations to make accurate predictions of user's behaviour. The other technique, conceptual drift, reflects the change over time of the attributes, that characterize a user. As a result of this change over time, historical models become invalid or insufficient. Finally, the last challange, efficiency issues covers large amount of data manipulation at every stage of user modelling and delivering results in real-time at prediction generation and on-line adaptation stages. These and other challenges are addressed in [66].

[71] studied on methods and techniques on adaptive hypermedia and states two different ways in learner modelling. One is collaborative approach, where user provides explicit feedback like questionnaires for determining learning styles. The other one is automatic learner modelling approach based on the behaviour and actions of the learner.

On the other hand, studies concentrated on learner modelling based on dominantly behavioural factors, are mostly semantic web based applications. [48] explains personalization in distributed e-learning networks as a part of Elena project. The proposed approach makes use of semantic web technologies and based on rule based matching of learning objects and learner descriptions. This rule-based personalization provides recommendation of related learning services or learning objects to appropriate learners. In the work [70], learner modelling is achieved just based on learner acts on a Web page. Further, in this work, explicit and implicit feedback is used to obtain Bayesian adaptive learner profiling. Another work regarding personalized e-learning in the semantic web by [60] proposes an approach based on ontologies for dynamically generating personalized hypertext relations powered by reasoning mechanisms. Additionally, not limited to e-learning, [65] concentrates to bring the ability to a search engine to automatically learn a user's preference based on his past click history and then to use the user preference to personalize search results. [65] shows that even from little click-history data user's preferences can be learned accurately and personalized search based on user preferences results in higher performance than existing ranking mechanism in the literature.

Further, in [47],[53],[54], learner modelling is based on ontology usage in different views to concentrate on a combination of personality, knowledge and behavioural factors. [54] proposes a learner modelling obtained using ontological abstraction. This ontological abstraction is used to construct a knowledge base in order to introduce the knowledge representations of the domain model and the curricular model. These knowledge representations are used in representing the prior knowledge of a learner and also assessment and continuously update of his knowledge [54]. Another work [47] is about interactive ontology-based learner modelling for personalized learning content management. The proposed approach integrates two existing systems, namely AIMS (task based information retrieval environment) and STyLE-OLM (interactive open learner modelling tool) in order to overcome some open issues regarding the adaptive learning content delivery. These mentioned open issues can be summarized as follows: Dealing with the empty learner model of first time learners, making use of learner's previous knowledge, explicit mechanisms to clarify the obtained learner models, compensating semantics and goal differences between teacher and learner [47]. [53] propose a student model integrated with ontology, enabling the personalization system to guide the system's learning process. This developed model monitors the student's progress so that it can update the concepts known by the student and decides which concepts s/he should learn next.

[44] describes a powerful learning management system (dotLRN), which is world-wide open source software for supporting e-learning and digital communities. dotLRN provides learner modelling, which reflect the gathered data from learners including their learning styles using Felder test, IMS-QTI questionnaires results, and also their active and passive interactions [44].

Regarding learning styles as a part of personality factors used in learner modelling, [63] concentrated on the critical issues influencing the design of adaptation based on the learning style information in Adaptive Educational Hypermedia Systems. Further, adaptive systems such as MASPLANG [64] and CS383 [45] detect learning styles using questionnaires.

[46] proposes an intelligent learning system with a specific user interface based on the Felder-Silverman learning style model. Using this interface, learning styles are detected from learner behaviour patterns on the interface using decision tree approach.

[51] evaluated Bayesian networks at detecting the learning style of a student in a Web-based education system based on his behaviour.

Furthermore, [55]-[58] contribute literature regarding detecting learning styles of learners in LMS based on Felder-Silvermann learning style model. [55] proposes an approach to detect learning styles in LMS based on the behaviour of learners during an online course. For learning styles Felder-Silverman learning style model is used .[56] provides additionally a practical example by extending the open-source Learning management system Moodle [72] with the proposed learning style detector tool. Further, [58] proposes an automatic student modelling approach for learning management systems (LMSs) for detecting learning style preferences according to the Felder-Silverman learning style model [30]. The proposed approach is based on patterns related to those features, implemented in most LMSs and used by teachers and course developers like content objects, outlines, self-assesment tests, exercises, examples and discussion forums. Then, the relevance of these patterns to each dimension of learning style is described and using a simple rule-based method students' behaviour is related to their preference of semantic groups [58]. In [57], an approach for profiling learners based on the answers to Index of Learning Style questionnaire (ILS) is presented. The proposed approach focuses on overcoming the limitations of the ILS questionnaire with respect to reliability and validity. The approach uses Multiple Correspondance Analysis and proximity measure to detect the learning style of the learner.

On the other hand, [52] discusses how learning styles information can be used in the design of a computer-based programming environment.

4.2. Learning Styles

Learning styles show the way of learning, which a learner may prefer. Various studies in the literature show the importance of learning styles in learning. In [24], Marton indicates the potential educational importance of information about students' learning styles, in higher education. The outcome of his studies brings a conclusion, that there are definite relationships between the ways in which an individual conceptualizes learning, the processes by which the individual attempts to learn, and the outcomes of the individual's attempts to learn [25]. Hence, in this approach, we use learning styles as our categories of users. For this purpose, first, we investigated learning styles in the literature in general. Then, we focused on our research on learning styles, which can be especially used in engineering science education domain, where diagrams, experiments, exercises, problem statements, figures, tables, graphs, narrative texts, lectures, slides, simulations are used as learning resource type. In this chapter, we present our literature survey results.

4.2.1. Background On Learning Styles

There are several learning style models reported in the literature. According to the report of learning & skills research centre [26], families of learning styles can be categorized into 5 categories as given below. The bold written phrases show the keywords for each of these learning style families.

1) Learning styles and preferences are largely **constitutionally-based** including the four modalities : VAKT (Visual, auditory, kinesthetic, tactile)
2) Learning styles reflect deep-seated features of the **cognitive structure,** including 'patterns of ability'.
3) Learning styles are one component of a relatively **stable personality type**.
4) Learning styles are **flexibly stable learning preferences.**
5) Move on from learning styles to **learning approaches, strategies, orientations and conceptions of learning.**

After introducing learning style families, we describe each of these families and give the names of learning style models falling into these families as follows [26]:

1) The theorists of learning styles falling into this category are claiming that learning styles are fixed or at least very difficult to change. To defend their claim, theorists refer to genetically influenced personality traits, or to the dominance of particular sensory or perceptual channels or to the dominance of certain functions linked with the left or right halves of the brain [25]. Learning style models falling into this category are namely as follows: Dunn and Dunn, Gregory, Bartlett, Betts, Gordon, Marks, Paivio, Richardson, Sheehan, Torrance

2) The theorists of learning styles of this family are considered to have a shared view of learning styles as 'structural properties of the cognitive system itself' [27]. Furthermore, they concentrate on the interactions of cognitive controls and cognitive processes. According to theorists in this family, styles are linked to particular personality features, with the implication that cognitive styles deeply embedded in personality structure. So, many of these styles are very similar to measures of ability. Learning style models falling into this category are namely as follows: Riding, Broverman, Cooper, and Gardiner et. al., Guilford, Holzman and Klein Hudson, Hunt, Kagan, Kogan, Messick, Pettigrew, Witkin.

3) The common feature of the learning style model in this family is that they concentrated on learning style concept as one part of the observable expression of a relatively stable personality type. The theorists of this family are concerned with constructing instruments which embed learning styles within an understanding of the personality traits that shape all aspects of an individual's interaction with the world. Learning style models falling into this category are namely as follows: Apter, Jackson, Myers-Briggs, Epstein and Meier, Harrison-Branson, Miler.

4) According to theorists of this family, learning style is not a fixed trait, but a differential preference for learning, which changes slightly from situation to situation. At the same time, there is some long term stability in learning style. Learning style models falling into this category are namely as follows: Allinson and Hayes, Herrmann, Honey and Mumford, Kolb, Felder and Silverman, Hermanussen, Wierstra, de Jong and Thijssen, Kaufmann, Kirton, McCarthy

5) Researchers within this family refer to underlying personality differences and relatively fixed cognitive characteristics. This leads them to differentiate between styles, strategies and approaches, with the latter being derived from perceptions of a task and cognitive strategies that learners might then adopt to tackle it. In this way, theorists of this family tend to avoid 'styles' in favor of strategies and 'approaches', because previous ideas about styles promoted the idea of specific interventions either to 'match' existing styles or to encourage a repertoire of styles. Learning style models falling into this category are namely as follows: Entwistle, Sternberg, Vermunt, Biggs, Conti and Kolody, Grasha-Riechmann, Hill, Marton and Saljö, McKenney and Keen, Pask, Pintrich, Smith, Garcia and McCeachie, Schmeck, Weinstein, Zimmerman and Palmer, Whetton and Cameron.

Before continuing with learning style models subject to engineering education, it should be mentioned that the concept of learning styles is not universally accepted. Some members of academic community argue that, learning style models have no sound theoretical basis and that the instruments used to asses learning styles have not been appropriately validated. On the other hand, instruction designed to address a broad spectrum of learning styles has consistently proved to be more effective than traditional instruction, which focuses on a narrow range of styles [28]. We therefore propose taking an engineering approach to learning styles, regarding them as useful heuristics for understanding students and designing effective instruction, and continuing to use them until demonstrably better heuristics appear [28].

4.2.2. Learning Style Models In Engineering Science Education

There are mainly 5 learning style models, which are subject of studies in the engineering science education literature [28]. This are namely Myers-Briggs Type Indicator (MBTI) [28],[29], Kolb's model [28],[31], Felder and Silverman model [28],[30], Herrmann Brain Dominance Instrument (HBDI) [28],[32] and Dunn and Dunn model [20],[33]. We are using Felder-Silverman model in the proposed automatic learner model module as presented in chapter 6.Therefore we will investigate it more closely, whereas we will briefly discuss the other models.

MBTI can be used to identify the personality type of a person and is probably one of the best-known instruments to assess learning styles. The MBTI is based on Carl Jung's theory of psychological types, which was extended and applied by Katherine Briggs and Isabel Briggs Myers. The indicator measures a person's preferences on four scales. Each preference may be mild or strong. The four scales form sixteen different psychological types.

The 4 scales are as follows [26]:

1) Direction of energy or interests

E	Extroversion	(outer world of people, try things out)
I	Introversion	(inner world of ideas and actions, think things through)

2) Preference of perception

S	Sensing	(immediate and practical experience, details, facts, procedures)
N	Intuition	(possibilities, meanings, imagination, concepts)

3) Preferences for decision making

T	Thinking	(logical, objective, rule-based, skeptical)
F	Feeling	(subjective, personal, value-based, appreciative)

4) Orientation to outside world

J	Judgement	(ordered, planned)
P	Perception	(spontaneous, adaptive)

Kolb's Learning Style Inventory (LSI) [35] assesses a student's learning abilities on two scales as illustrated in Figure 4.1. The first scale applies to the way people grasp experience. On one end this can be through experiencing the concrete, tangible, felt qualities of the world. Others tend to perceive, grasp, or take hold of new information through symbolic representation or abstract conceptualization. They think about, analyze and systematically plan things. The second scale applies to how people transform or process experience. Some tend to carefully watch others who are involved in the experience and reflect on what happens (reflective observation), while others choose to jump right in and start doing things (active experimentation). From these two scales the four learning styles of the LSI are defined as: diverging (dromer), assimilating (denker), converging (beslisser), and accommodating (doener).

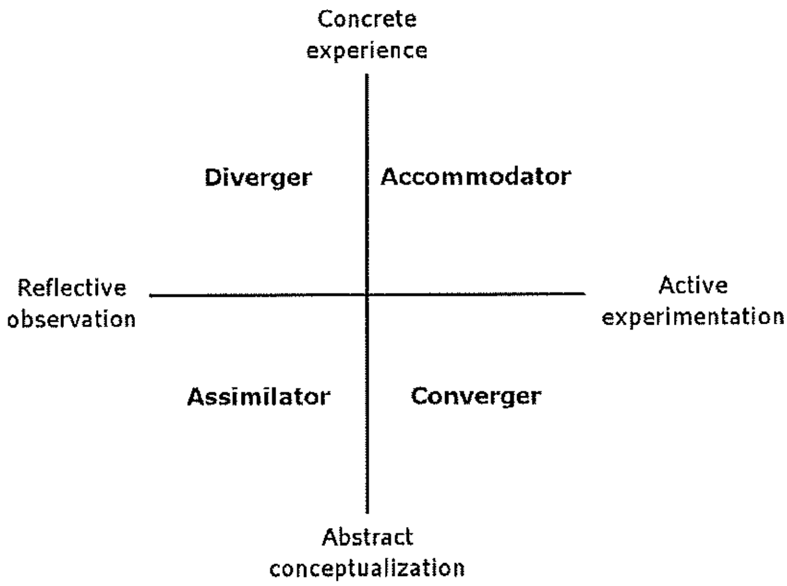


Figure 4.1 Kolb's four learning styles ([31])

HBDI was constructed by Herrmann [32],[36]. It classifies students in terms of their relative preferences for thinking in four different modes based on the task-specialized functioning of the physical brain, as shown in Figure 4.2.

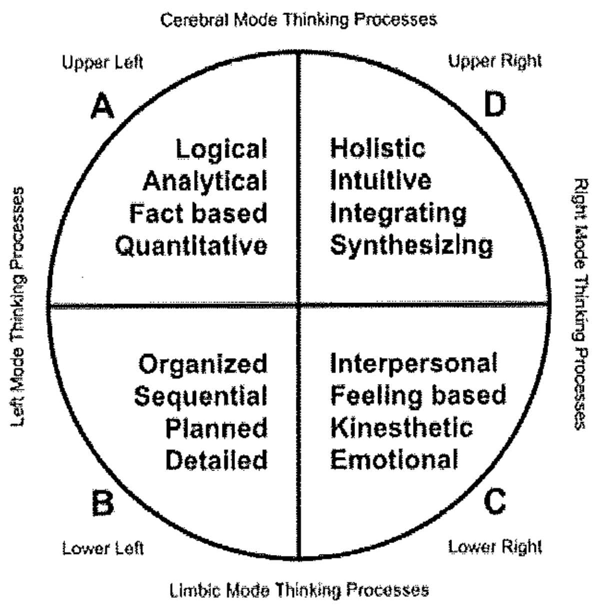


Figure 4.2 The Whole Brain model ([37])

According to the Dunn and Dunn model [26], learning style is divided into 5 major strands called *stimuli.* The stimulus strands are: a) environmental, b) emotional, c) sociological, d) psychological, and e) physiological elements that significantly influence how many individuals learn [33]. From these strands, four variables affect students' preferences, each of which includes different factors. These are measured in the model and summarized in Table 4.1.

Table 4.1 Variables with different factors affecting student's preferences according Dunn and Dunn model ([26]).

Variable	Factors			
Environmental	Sound	Temperature	Light	Seating, layout of room, etc
Emotional	Motivation	Degree of responsibility	Persistence	Need for structure
Physical	Modality preferences – ie for visual, auditory, kinaesthetic or tactile learning (VAKT)	Intake (food and drink)	Time of day	Mobility
Sociological	Learning groups	Help/support from authority figures	Working alone or with peers	Motivation from parent/teacher

4.2.2.1. Felder-Silverman Learning Style Model

This learning style model was introduced by Felder and Silverman [30] and was inspired by experiences in engineering education. It has 4 dimensions to show preferred learning style of a student. These dimensions are as follows [38]:

1) Preferred perceived type of information:

<u>Sensory (sights, sounds, physical sensations):</u> Sensory learners are concrete, practical, methodical, and oriented toward facts and hands-on procedures.

Intuitive (memories, thought, insights): Intuitive learners are more comfortable with abstractions (theories, mathematical models) and are more likely to be rapid and innovative problem solvers.

2) Most effective input channel:

Visual (pictures, diagrams, flow charts, demonstrations): Visual learners prefer visual representations of presented material

Verbal: Verbal learners prefer written and spoken explanations

3) Preferred processing of information:

Active (through engagement in physical activity or discussion): Active learners learn by trying things out, working with others.

Reflective (through introspection): Reflective learners learn by thinking things through, working alone.

4) Progress to understanding:

Sequential (in a logical progression of incremental steps): Sequential learners tend to think in a linear manner and are able to function with only partial understanding of material they have been taught.

Global (in large "big picture" jumps): Global learners_think in a systems-oriented manner, and may have trouble applying new material until they fully understand it and see how it relates to material they already know about and understand. Once they grasp the big picture, however, their holistic perspective enables them to see innovative solutions to problems that sequential learners might take much longer to reach, if they get there at all.

Index of Learning Styles (ILS): The Index of Learning Styles® (ILS) is a forty-four-item forced-choice instrument developed in 1991 by Richard Felder and Barbara Soloman to assess preferences on the four dimensions of the Felder-Silverman model. The ILS is available at no cost to individuals who wish to assess their own preferences and to instructors or students who wish to use it for classroom instruction or research, and it may be licensed by non-educational organizations. For a closer look at ILS and its result, an example is presented in Appendix-B. Further, the details of ILS questionnaire evaluation mechanism are presented in 6.5. Experimental Results.

Regarding the proper use of learning styles and the instruments to assess them, [39] make the following statements:

- Learning style dimensions are continua, not either/or categories. A preference for one pole of a dimension may be mild, moderate or strong.
- Learning style profiles suggest behavioral tendencies rather than being infallible predictors of behavior.
- Learning style preferences are not reliable indicators of learning strengths and weaknesses.
- Learning style preferences can be affected by a student's educational experiences.
- The point of identifying learning styles is not to label individual students and modify instruction to fit their labels. The optimal teaching style is a balanced one in which all students are sometimes taught in a manner that matches their learning style preferences, so they are not too uncomfortable to learn effectively, and sometimes in the opposite manner, so they are forced to stretch and grow in direction they might be inclined to avoid if given the option. (Note that this statement is made with engineering education in mind.)

4.2.2.2. Parallelism Between Models

There exist several parallels between the Felder–Silverman model and other learning style models described above, some of which are:

- The active/reflective dimension is similar to the same dimension of Kolb's LSI and relates to the extrovert/introvert dimension of the MBTI.
- The sensing/intuitive dimension is the same as the one of the MBTI and may match the concrete/abstract dimension of Kolb's model.

- The visual/verbal dimension has connections with the visual–auditory–kinesthetic formulation of modality theory, neurolinguistic programming and cognitive studies of information processing.
- The sequential/global dimension is alike left–brain/right–brain dominance modes.

4.2.2.3. Felder–Silverman Learning Style Model In Personalization

In literature, there are various works available providing personalization issue in e-learning by diagnosing learning styles of users based on the Felder-Silverman learning style model. These works can be stated as follows:

[44] describes a powerful learning management system (dotLRN), which is world-wide open source software for supporting e-learning and digital communities. dotLRN provides learner modelling, which reflect the gathered data from learners including their learning styles using Felder test, IMS-QTI questionnaires results, and also their active and passive interactions [44].

[46] proposes an intelligent learning system with a specific user interface based on the Felder-Silverman learning style model. Using this interface, learning styles are detected from learner behaviour patterns on the interface using decision tree approach.

[51] evaluated Bayesian networks at detecting the learning style of a student in a Web-based education system using his behaviour based on the Felder-Silverman learning style model.

Furthermore, [55]-[58] contribute literature regarding detecting learning styles of learners in LMS based on Felder-Silvermann learning style model. [55] proposes an approach to detect learning styles in LMS based on the behaviour of learners during an online course. For learning styles Felder-Silverman learning style model is used

.[56] provides additionally a practical example by extending the open-source Learning management system Moodle [72] with the proposed learning style detector tool. Further, [58] proposes an automatic student modelling approach for learning management systems (LMSs) for detecting learning style preferences according to the Felder-Silverman learning style model [30]. The proposed approach is based on patterns related to those features, implemented in most LMSs and used by teachers and course developers like content objects, outlines, self-assesment tests, exercises, examples and discussion forums. Then, the relevance of these patterns to each dimension of learning style is described and using a simple rule-based method students' behaviour is related to their preference of semantic groups [58]. In [57], an approach for profiling learners based on the answers to Index of Learning Style (ILS) questionnaire is presented. The proposed approach focuses on overcoming the limitations of the ILS questionnaire with respect to reliability and validity. The approach uses Multiple Correspondance Analysis and proximity measure to detect the learning style of the learner.

CHAPTER 5

CLASSIFICATION ALGORITHMS

In e-learning systems, where peronalization is provided via learner models, classification is used in learner model extraction. In this chapter, we briefly summarize classification algorithms for textual data, because learner query results will be in textual data format.

More explicitly, we are addressing the problem of how to classify an arbitrarily sized database of labeled instances, where attributes are not necessarily independent. In the literature, existing solutions for this problem [75] are Naive-Bayes classifiers, decision trees (C4.5) and a hybrid solution NBTrees.

Therefore, in this chapter, although there are various classification algorithms available for text categorization like Rocchio's , kNN, ANN, Voting algorithms and SVM, we briefly explain the Naive-Bayes classsifiers,decision trees (C4.5) and NBTree algorithms, that can be used for the problem that we addressed with their advantages and disadvantages [40]-[43], [75], [84],[85].

Furthermore, as a requirement of the proposed automatic learner modeling approach, multi-label classification is mentioned in this section. The requirement and usage of multi-label classification in this thesis is mentioned in section 6.

5.1. Decision Tree (DT) Algorithm

ID3 and C4.5 are the basic algorithms for inducing decision trees. C4.5 is an extension of ID3, that covers also unavailable values, continuous attribute value ranges, pruning of decision trees, rule derivation, and so on.In decision tree agorithm, documents are organized in a tree form, such that every leaf node represents a class and seperation into subsets from root is done according to an attribute. This algorithm is easy to understand and to generate rules.It is fast and provides an easy interpretable segmentation of data. Also, it decreases problem complexity,whereas training time is relatiely expensive and a document is only connected with one branch.Therefore, it may suffer from overfitting and does not handle continuous variable well.Further, interpretability decreases when the number of splits increases.

5.2. Naive Bayes (NB) Algorithm

This algorithm estimates the probability of each class for a document using Baye's rule. It works well for numeric and textual data and easy to implement and calculate.It is fast and based on very strong independence between attributes assumption. It uses evidence from many attributes and at the same time robust to irrelevant attributes.. However, conditional independence assumption is not applicable in real-world and performs very poorly when features are highly correlated or database is large.

5.3 NBTree Algorithm

NBTree algorithm is a hybrid algorithm, which causes a hybrid of Naive-Bayes classifiers with decision tree classifiers. This algorithm is similar to the classical decision trees except that leave nodes are Naive-Bayes categorizers instead of nodes predicting a single class. NBTree algorithm is appropriate when many attributes are relevant for classification and attributes are not necessarily independent. Moreover, it provides easy interpretability and is suitable for large databases outperforming better than decision trees and NB classifiers alone[75] .

5.4. Multi-Label Classification

In some classifcation tasks, it is likely that some data belongs to multiple classes, causing the actual classes to overlap by definition. In text, music or scene categorization, documents may belong to multiple genres, such as government and health, or rock and blues [86] or a photograph can belong to more than one conceptual class, such as *sunsets* and *beaches* at the same time.

Also, in our proposed learner modeling approach, which will be presented in chapter 6 in detail, learners with different learning characteristics can have common data object selections. This situation addresses multi-label classification. Therefore, in this section, we present different multi-label classification methods available in the literature and introduce evaluation metrics for them.

Formally, in multi-label classification, samples are associated with a set of labels $Y \subseteq L$. An example for a multi-label data set is given in Table 5.1. In this table, there are 4 document samples, each of them belong to one or more of four classes: Health, Government, Education, and Economy.

Table 5.1 An example for a multi-label data set

Sample	Health	Government	Education	Economy
1	X			X
2			X	X
3	X			
4		X	X	

In [87], multi-label classification methods are categorized into two main categories: a) problem transformation methods, and b) algorithm adaptation methods.

Problem transformation methods transform the multi-label classification problem either into one or more single-label classification or regression problems [87].

On the other hand, algorithm adaptation methods extend specific learning algorithms in order to handle multi-label data directly [87].

The data set of Table 5.1 is used to exemplify the methods in the following sections.

5.4.1. Problem Transformation Methods

According to [87], there are two simple methods to convert multi-label classification problem into single-label classification. One method for this conversion is randomly or subjectively selecting one of the multiple labels for each of the multiple-labelled samples as presented in Table 5.2 and called here as method-1. The other method, called here method-2, is just ignoring multi-labelled samples from the data set and transformed data set is as illustrated in Table 5.3.

Table 5.2 Transformed data set using method-1

Sample	Health	Government	Education	Economy
1	X			
2				X
3	X			
4			X	

Table 5.3 Transformed data set using method – 2

Sample	Health	Government	Education	Economy
3	X			

In both methods, a majority part of the information content of multi-label data set is lost. Therefore, these methods are inefficient to be used.

In an another method, called method-3, investigating multi-labelled samples, each different set of labels constructs a new single label for the multi-label data set as shown in Table 5.4. This method known in literature [86], [87] as one single-label classifier: H→P (L), where P (L) is the power set of L. So method-3 uses Label Powerset (LP) Classifier. The problem with this method is that, it can cause large number of classes and sparse data distribution among these multiple classes to build usable models. This method is used in literature works [86], [88].

Table 5.4 Transformed data set using method – 3 (LP)

Sample	Health	(Health&Economy)	(Government&Education)	(Education&Economy)
1		X		
2				X
3	X			
4			X	

Furthermore, method-4 is a common method and like dealing with a single-label multi-class problem using a binary classifier. This method learns $|L|$ binary classifiers Hl: $X \rightarrow \{l, \neg l\}$, one for each different label l in L. It transforms the original data set into $|L|$ data sets Dl that contain all examples of the original data set, labelled as l if the labels of the original example contained l and as $\neg l$ otherwise [87]. Hence, method-4 uses Binary Relevance (BR) Classifier. Accordingly, the multi-label data set given in Table 4.2 is converted to 4 data sets in this method as presented in Table 5.5.

Table 5.5 Transformed data set using method – 4

Sample	Health	¬ Health
1	X	
2		X
3	X	
4		X

Sample	Government	¬ Government
1		X
2		X
3		X
4	X	

Sample	Education	¬ Education
1		X
2	X	
3		X
4	X	

Sample	Economy	¬ Economy
1		X
2		X
3	X	
4	X	

With this method, in classifying a new instance x, output will be a set of labels. This method is used in literature works [86], [89]-[91].

Another method, called method -5 [87] decomposes each example (x, Y) into $|Y|$ examples (x, l) for all $l \in Y$. Then it learns one single-label *coverage-based* classifier from the transformed data set. Distribution classifiers are those classifiers that can output a distribution of certainty degrees (or probabilities) for all labels in L. Finally it post-processes this distribution to output a set of labels. One simple way to achieve this is to output those labels for which the certainty degree is greater than a specific threshold (e.g. 0.5) or a percentage (e.g. 70%) of the highest certainty degree [87]. The transformed data set is illustrated in Table 5.6.

Table 5.6 Transformed data set using method – 5

Sample	Class
1	Health
1	Economy
2	Education
2	Economy
3	Health
4	Government
4	Education

5.4.2. Algorithm Adaptation Methods

In [92], the C4.5 algorithm is adapted for multi-label data by modifying the formula of entropy calculation as follows:

$$Entropy(S) = -\sum_{i=1}^{N}(p(c_i)\log p(c_i) + q(c_i)\log q(c_i)) \qquad (5.1)$$

Where $p\ (c_i)$ = relative frequency of class c_i and $q\ (c_i) = 1-p\ (c_i)$. They also allowed multiple labels in the leaves of the tree.

Also, AdaBoost [93] is extended as presented in [94] to apply AdaBoost on weak classifiers.

Although these two algorithms are adaptations of a specific learning approach, at their core, they actually use method-6: Each example (x, Y) is decomposed into $|L|$ examples $(x, l, Y[l])$, for all $l \in L$, where $Y[l] = 1$ if $l \in Y$, and $[l] = -1$ otherwise. Table 5.7 shows the transformed data set obtained using method - 6.

Table 5.7 Transformed data set using method – 6

Sample	l	Y[l]
1	Health	1
1	Government	-1
1	Education	-1
1	Economy	1
2	Health	-1
2	Government	-1
2	Education	1
2	Economy	1
3	Health	1
3	Government	-1
3	Education	-1
3	Economy	-1
4	Health	-1
4	Government	1
4	Education	1
4	Economy	-1

ML-kNN [95] is an adaptation of kNN lazy learning algorithm for multi-label data following the paradigm of method -4.

In [96], a probabilistic generative model is defined for producing multi-label document and uses Bayes rule by following the paradigm of method-3.

[97] present a ranking algorithm for multi-label classification based on SVM. Furthermore, in [97] SVM is used with method-4 for multi-label classification problem.

Also, [99] is an algorithm that follows the paradigm of associative classification, which deals with the construction of classification rule sets using association rule mining.

5.4.3. Evaluation Metrics

Multi-label classification requires different evaluation metrics than those used in traditional single-label classification [87]. Before presenting these evaluation metrics from literature, we have to mention some other concepts. These concepts, namely label cardinality and label density, influence the performance of the different multi-label methods.

Let D be a multi-label data set consisting of |D| multi-label examples (xi, Yi), i = 1...$|D|$ and $Y_i \subseteq L$.

Label cardinality of D is the average number of labels of the examples in D:

$$LC(D) = \frac{1}{|D|}\sum_{i=1}^{|D|}|Y_i| \tag{5.2}$$

Label cardinality is independent of the number of labels $|L|$ in the classification problem, and is used to quantify the number of alternative labels that characterize the examples of a multi-label training data set.

Label density of D is the average number of labels of the examples in D divided by $|L|$:

$$LC(D) = \frac{1}{|D|}\sum_{i=1}^{|D|}\frac{|Y_i|}{|L|} \tag{5.3}$$

Label density takes into consideration the number of labels in the classification problem.

As a result, two data sets with the same label cardinality but with a great difference in the number of labels (different label density) might not exhibit the same properties and cause different behaviour to the multi-label classification methods. The two metrics are related to each other: LC (D) = $|L|$ LD (D).

Regarding evaluation metrics,

Let H be a multi-label classifier and $Z_i = H(x_i)$ be the set of labels predicted by H for example x_i.

$$\text{Hamming Loss}(H,\Delta) = \frac{1}{|D|}\sum_{i=1}^{|D|}\frac{|Y_i \Delta Z_i|}{|L|} \tag{5.4}$$

where Δ stands for the symmetric difference of two sets and corresponds to the XOR operation in Boolean logic.

The following metrics are used in [98] for the evaluation of H on D:

$$\text{Accuracy}(H,D) = \frac{1}{|D|}\sum_{i=1}^{|D|}\left|\frac{Y_i \cap Z_i}{Y_i \cup Z_i}\right| \tag{5.5}$$

$$\text{Precision}(H,D) = \frac{1}{|D|}\sum_{i=1}^{|D|}\left|\frac{Y_i \cap Z_i}{Z_i}\right| \tag{5.6}$$

$$\text{Recall}(H,D) = \frac{1}{|D|}\sum_{i=1}^{|D|}\left|\frac{Y_i \cap Z_i}{Y_i}\right| \tag{5.7}$$

Furthermore, in label based evaluation, to evaluate the metrics like accuracy, precision and recall, two operations for averaging across all labels can be used. These are microaveraging and macroaveraging.

Microaveraging is done by collecting decisions for all classes, computing contingency table and evaluating it, whereas macroaveraging is done by computing performance for each class, then averaging it.

CHAPTER 6

PROPOSED AUTOMATIC LEARNER MODEL MODULE

For the personalization part in an e-learning system, we introduce our proposed automatic learner model module. First, we present the design of the proposed module by explaining its subparts and their working principles and implementation details individually. Then, after explaining the selected reference instrument for the evaluation of the correctness of identifying learning styles at the output of the proposed module in identifying, we present the experimental results.

Automatic learner model module represents our contribution to personalization in e-learning system based on learning styles. Therefore, we present our proposed learner model for e-learning, which is derived to automatically infer learning style from learner's most interested data objects instead of using learning style assessment tools like questionnaires, which require explicit feedback from the learner. As the learning style model reference, we used Felder-Silverman learning style model, because it seems the most appropriate one for computer-based educational systems [45]. Further, we used NBTree algorithm in conjunction with Binary Relevance classifier, because it is the most performance-rich algorithm according to multi-label evaluation metric values presented in Table 6.4 and Table 6.5 among the investigated algorithms for our case, where there are a small set of attribute-value pairs and attributes take only a small set of distinct possible values and are non mutually exclusive. Hence, we can use the advantages of both decision trees and Naïve Byes in a hybrid form. The general overview of our proposed automatic learner model, designed by ourselves, is illustrated in Figure 6.1.

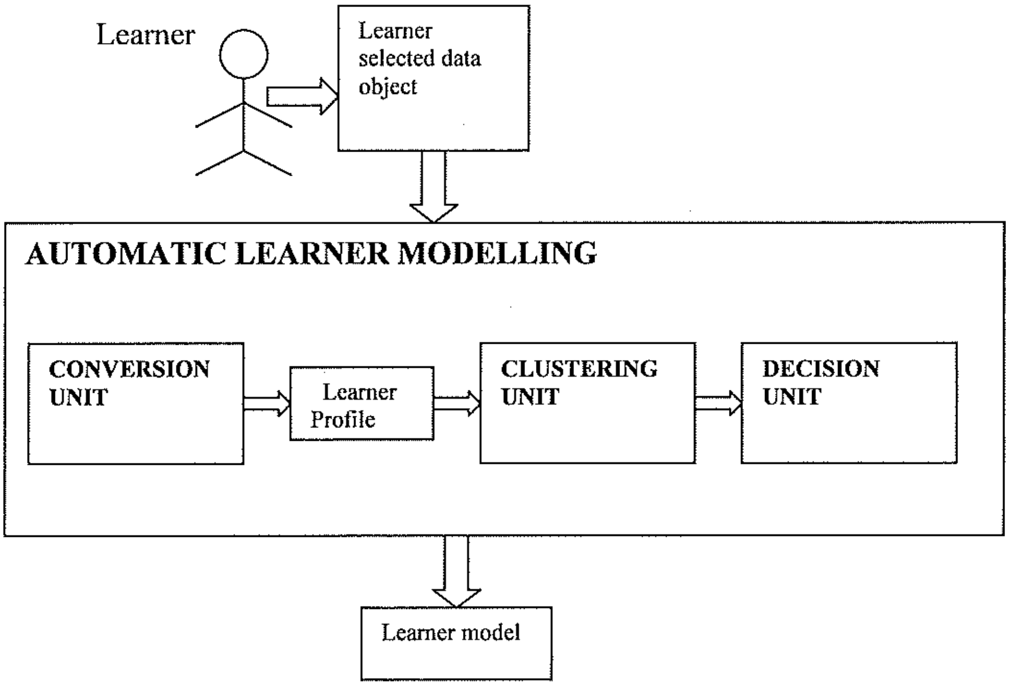


Figure 6.1 The General Overview of Proposed Automatic Learner Model Module

The learner first submits a generic query for the topic he is interested in. From the returned data, the learner selects the ones that most fulfills his requirements. We will name these selections as the learner selected data objects (LSDOs) from now on. Then, using the conversion unit given in Figure 6.2, learner profile table is obtained. This table is then processed by the clustering unit which is based on NBTree classification used in conjunction with Binary Relevance classifier. The training of the system for the clustering unit has to be done using a training data set reflecting different learning styles as an offline process.

The clustering unit assigns labels to each row of the learner profile table using the dimensions of the FSLSM. After that, using the decision unit, learning style is detected and hence learner model is obtained using the dominant pole in each dimension of the FSLSM.

The above process is done for each learner only at their first log on to the system as an offline process. Then, in the consecutive logons of the same learner, his learner model based on his learning style is used to return the learner most effective and suitable results for his submitted generic query. Moreover, taking into account that, learning style behavior of the learner changes slowly along the time, using the proposed approach with certain time intervals, the changes in the learner's learning characteristics over time can be followed. Hence, a dynamically updated learner model can be obtained.

In the following subsections, we will explain each part of the proposed system in detail.

6.1. Profile Table

The profile table consists of four columns corresponding to attributes, i.e. the metadata names. The attributes "LearningResourceType" and "RequiredDegreeOfCollaboration" are [2] metadata fields and the other two attributes are extended fields. These fields are added to derive characteristics representing each dimension of the selected learning style model. Each row of the table takes predetermined values of these attributes given in Table 6.1. Further, the whole profile table is presented in Appendix –C.

Table 6.1 Predetermined attribute values

NAME OF ATTRIBUTE	POSSIBLE PREDETERMINED VALUE SET
LEARNINGRESOURCETYPE (LRT)	EXERCISE / SIMULATION-DEMO / VISUAL RESOURCES / EXPERIMENT / WRITTEN TEXT / LECTURE
REQUIREDDEGREEOF COLLABORATION (RDOC)	INDIVIDUAL WORK / TEAM WORK / WITH THE ASSISTANCE OF THE TEACHER
TYPEOFCOVERAGE (TOC)	THEORETICAL-CONCEPTS, DISCOVER RELATIONSHIPS / PRACTICAL-KNOWN METHODS, APPLICABLE TO REAL WORLD
THEARTOFPRESENTING CONTENT (TAOPC)	SEQUENTIALLY, STEP BY STEP / GLOBALLY-WITH THE BIG PICTURE OF SUBJECT

The length of this table is constant such that it covers all the possible combinations of the attribute values. Each learner profile table can be considered as a sub part of this profile table.

6.2. Conversion Unit

The aim of the conversion unit is to construct the learner profile table from LSDOs. For this only LSDOs are used rather than learner behavior observed in time scale. The reason is that to be able to propose an automatic learner modeling approach with independency from the underlying background like LMSs.

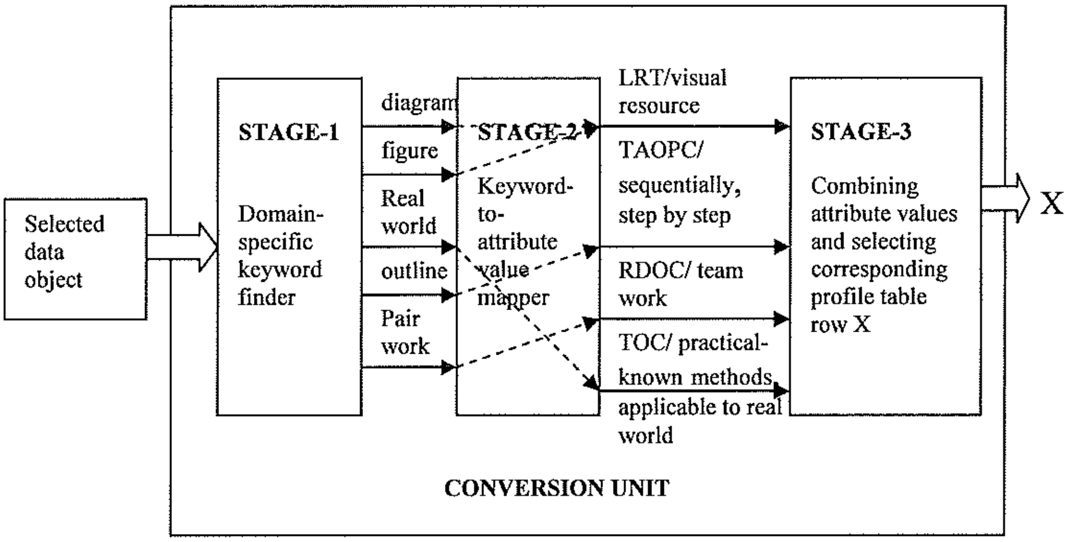
This conversion unit is domain-specific and composed of three stages: (1) Domain-specific keyword finder, (2) keywords-to-attribute-value mapper, (3) learner profile table construction. The keyword finder is based on the dominant characteristics of each dimension ([28], [30]), the semantic grouping of each dimension ([105]) and the LSDO properties (containing mostly drawings or textual data, subsections etc.) The keywords presented on Table 6.2 can be extended based on a further detailed pedagogical and cognitive study.

Table 6.2 Relevance of keywords for groups of learning styles

DIMENSION NAME OF LEARNING STYLE MODEL	RELEVANT KEYWORDS
GLOBAL (G)	OVERALL, OVERVIEW, TOTALLY,ABSTRACT,WHOLE
SEUENTIAL (SEQ)	SEQUENTIAL, OUTLINE, FIRST, SECOND, FLOWCHART, DETAIL, SUBSECTIONS,PARTS
ACTIVE (A)	EXPERIMENTALLY, PAIR WORK,USUALLY,ORDINARY
REFLECTIVE (R)	OBSERVATION, THEORY, THEOREM,CHALLENGES,ALONE WORK
INTUITIVE (I)	THEORETICALLY, IN PRINCIPAL
SENSING (S)	PRACTICALLY, IN REAL WORLD APPLICATIONS,EXPERIMENTAL DATA RESULTS
VISUAL(VIS)	FIGURE,GRAPH,CHART,*.JPG,*.GIF,*.BMAP,*.AVI, TABLE, DIAGRAM, SCHEMA,VIDEO
VERBAL(VER)	FORUM, DISCUSSION BOARD, TEXT

Once the keywords are found, they are mapped to the corresponding attribute values given in Table 6.2. Finally, the attribute values are combined to select the corresponding row in the profile table. The collection of these rows, form the learner profile table.

Learner profile table construction from LSDOs using the conversion unit is illustrated with an example in Figure 6.2. In Figure 6.2, at stage-1 domain-specific keywords are found. Then, at stage-2, these keywords are mapped to corresponding attribute values. After that, at stage-3, these attribute value pairs are combined and the corresponding learner profile table row is constructed.



X:

(LRT)	(RDOC)	(TOC)	(TAOPC)
VISUAL RESOURCE	TEAM WORK	PRACTICAL-KNOWN METHODS,	SEQUENTIALLY, STEP BY STEP

Figure 6.2 Conversion Unit

It should be noticed that, for learners with different learning style dimensions, the same profile table rows; i.e. the same combination of keywords can exist. This means, the attributes of learner profile rows are dependent. The same attribute value can be used in evaluating different learning style dimensions. Hence, we permit multi-labeling in the proposed approach. This is one of the differences of the proposed automatic learner modeling approach. In [46] and [51], the described relevant patterns for each learning style dimension are disjoint from each other. In other words, a pattern is not used for any two different learning style dimensions.

6.3. Clustering Unit

The aim of the clustering unit is to assign labels to each row of the learner profile table using the dimensions of the FSLSM. The clustering is done based on the predetermined classes obtained by NBTree classification in conjunction with Binary Relevance classifier from a training data set.

For preparing the training data set, we collect LSDOs from learners with different learning styles. Then using the conversion unit, we get the learner profile table of each learner. Furthermore, we get ILS results for each learner simultaneously. For the training data set, we label each row of learner profile table with the corresponding ILS result. Hence, the training data set is composed of correctly labeled rows. However, as stated previously learners with different learning styles can have common LSDOs causing two rows with the same content having different labels. As a result, we get multi-label classification input. In literature, to handle multiple labels, multi-label classification methods are used ([86], [95]).

These methods can mainly be grouped into two groups: transformation methods and algorithm adaptation methods ([87]). Problem transformation methods transform the multi-label classification problem either into one or more single-label classification or regression problems, whereas algorithm adaptation methods extend specific learning algorithms in order to handle multi-label data directly. The main problem transformation methods are Label Powerset (LP), Binary Relevance classifier (BR), problem transformation method-6 (PT-6) ([87]).

In LP, for each different set of labels in multi-labeled samples, a new single label is used.

BR method learns $|L|$ binary classifiers Hl: $X \rightarrow \{l, \neg l\}$, where l represent each different label in L, H represents a binary classifier and X represent a transformed multi-labeled sample. It transforms the original data set into $|L|$ data sets, where each of them contains all examples of the original data set, labelled as l if the labels of the original example contained l and as $\neg l$ otherwise.

In PT-6, each example x with a multi-label set Y is decomposed into $|L|$ examples (x, l, Y[l]), for all $l \in L$, where $Y[l] = 1$ if $l \in Y$, and $Y[l] = -1$ otherwise.

The last step of the classification algorithm is to classify the single-labeled rows. If the attributes of these labeled rows are not totally independent (which is the case in our problem), it is common to use NB or NBTree, a hybrid of NB classifiers and decision tree classifiers ([75]).NB algorithm estimates the probability of each class using Baye's rule. On the other hand, NBTree algorithm is a hybrid algorithm. It is similar to the classical decision trees except that leave nodes are Naive-Bayes categorizers instead of nodes predicting a single class. NBTree algorithm is appropriate when many attributes are relevant for classification and attributes are not necessarily independent. Regarding decision tree induction, ID3 and C4.5 are the basic algorithms, where C4.5 is an extension of ID3, that covers also unavailable values, continuous attribute value ranges, pruning of decision trees, rule derivation, and so on ([75]).

Hence, the proposed clustering unit is composed of a classification algorithm in conjunction with a multi-label classification method. In the literature, it is shown that different classification algorithms in conjunction with different multi-label classification methods illustrate different performance results depending on the experimental data properties ([87]). Due to this, in order to choose the most reliable classification for our domain, we evaluate different classification algorithms with different multi-label classification methods using cross-validation technique and get performance results based on the evaluation measures like hamming loss, precision, accuracy and recall, ([87]). Evaluation measures can be categorized into two groups: example based evaluation measures and label based evaluation measures ([106]). Example based evaluation measures (e.g. hamming loss, accuracy, precision, recall) use the difference between actual and predicted label sets for each example and average the results over all examples. Label based evaluation measures (e.g. accuracy, precision, recall) calculate a binary evaluation separately for each label and micro/macro average the results across all labels ([106]).

The example based evaluation measures used in the performance results are defined as presented by equations 5.4-5.7 in section 5.4.3.

Among the performance results, the one with maximum accuracy and precision, minimum hamming loss should be selected. In order to choose the algorithm for our clustering unit, we cluster our training data set using 3 different multi-label classification methods (LP,BR,PT-6) with four different classification algorithms (NBTree,NB,ID3,C4.5).

In obtaining the results, we used our training data set where the correct class labels are assigned to each row from the ILS results. The class labels represent 4 dimensions of the FSLSM. The class labels are presented with the following abbreviations: REF (Reflective), ACT (Active), INT (Intuitive), SEN (Sensing), VIS (Visual), VER (Verbal), SEQ (Sequential), GLO (Global), and NEU (Neutral/Balanced). NEU is used only if the learner is balanced for a dimension. For example, if the learner is strong on Reflective pole for processing dimension and strong in Visual pole of input dimension and balanced in the rest 2 dimensions, then this person is labeled as REF-VIS-NEU. In table 6.6, we showed all 65 different labels, covering all possible combinations of the given abbreviations. Also, the labels, which are available in our training set, are showed as bold italic.
The properties of our training data set are presented in Table 6.3. In Table 6.3, label cardinality represents the average number of labels of the multi-labeled rows and label density is label cardinality divided by number of labels. Different multi-label data sets with the same label cardinality but with the different label densities can cause different performance results in multi-label classification methods. Therefore, we presented the properties of the training data set.

Then, we used different multi-label classification methods on our training data set and obtained different transformed data sets for each multi-label classification method (LP, BR, and PT-6). After that, for each of the transformed data set we run different classification algorithms (ID3, C4.5, NB, NBTree) separately and get predicted label set. We evaluated the performances by the equations of section 5.4.3.using the predicted and assigned labels for each case and the results are shown in Table 6.4 and 6.5.

Table 6.4 presents performance results based on example based evaluation and Table 6.5 presents performance results based on label based evaluation using microaveraging. Microaveraging is done by collecting decisions for all labels, computing contingency table and evaluating it.

Evaluating Table 6.4 and Table 6.5, we get maximum accuracy for binary relevance classifier with NBTree classifier. Also, for this case, hamming loss is minimum and precision is maximum among all. As a result, in the proposed novel approach, using cross-validation with our training data set, we use NBTree classification algorithm in conjunction with binary relevance classifier inside the clustering unit.

Once the clustering unit is trained as described above as an offline process, at run time it works as follows for test data set:

First, LSDOs from learner with unknown learning style are converted to learner profile rows. Then, these learner profile rows are classified using the clustering unit and for each learner profile row the predicted label set is obtained.

Figure 6.3 Clustering Unit

Table 6.3 Properties of training set

No. Of Instances		Attributes		Labels	Label Density	Label Cardinality
Training set Before Multi-Labelling -with Overlapping Instances	Multi-labelled Training set	Numeric	Discrete			
244	68	0	4	10	0.359	3.59

Table 6.4 Performance results of training set with different candidate classification algorithms in conjunction with multi-label methods - with example based evaluation.

	method-3 (LP)				method – 4 (BR)			
Metric - Example-Based Evaluation	ID3	C4.5	NB	NB Tree	ID3	C4.5	NB	NB Tree
Hamming Loss	0.206	0.245	0.270	0.249	0.188	0.168	0.232	0.153
Accuracy	0.572	0.514	0.446	0.488	0.591	0.612	0.504	0.633
Precision	0.727	0.686	0.673	0.689	0.715	0.835	0.725	0.855
Recall	0.720	0.648	0.557	0.605	0.746	0.709	0.625	0.694

	method- 6			
Metric-Example-Based Evaluation	ID3	C4.5	NB	NB Tree
Hamming Loss	0.187	0.168	0.282	0.172
Accuracy	0.594	0.615	0.407	0.594
Precision	0.715	0.832	0.628	0.827
Recall	0.749	0.715	0.578	0.665

Table 6.5 Performance results of training set with different candidate classification algorithms in conjunction with multi-label methods - with label based evaluation (microaveraging)

	method-3 (LP)				method – 4 (BR)			
Metric- Label Based Evaluation (Microaveraging)	ID3	C4.5	NB	NB Tree	ID3	C4.5	NB	NB Tree
Accuracy	0.794	0.755	0.730	0.751	0.812	0.832	0.768	0.847
Precision	0.715	0.671	0.667	0.682	0.738	0.849	0.727	0.876
Recall	0.712	0.633	0.536	0.595	0.755	0.671	0.598	0.694

	method- 6			
Metric- Label Based Evaluation (Microaveraging)	ID3	C4.5	NB	NB Tree
Accuracy	0.813	0.832	0.718	0.828
Precision	0.738	0.845	0.623	0.851
Recall	0.758	0.679	0.592	0.645

The labels presented in Table 6.6 represent 4 dimensions of the Felder-Silverman learning style model. In label naming, for the strong pole of any dimension it is mentioned with the corresponding abbreviation and NEU is used if the learner is balanced for a dimension. For example, if the learner is balanced in all dimensions, then we label his samples as NEU. On the other hand, if the learner is strong on Reflective pole for processing dimension and strong in Visual pole of input dimension and balanced in the rest 2 dimensions, then this person is labeled as REF-VIS-NEU.

Table 6.6 Labels of the classes

Label Of The Class		
NEU	GLO-VIS-NEU	SEN-NEU
REF-NEU	ACT- NEU	***SEN-VIS-NEU***
INT-VIS-NEU	INT-VER-NEU	SEN-VER-NEU
REF-VIS-NEU	REF-VER-NEU	ACT-VER-NEU
SEQ-NEU	GLO-NEU	ACT-VIS-NEU
VIS-NEU	VER-NEU	INT-NEU
GLO-VER-NEU	SEQ-VIS-NEU	SEQ-VER-NEU
REF-SEQ-NEU	REF-GLO-NEU	ACT-GLO-NEU
ACT-SEQ-NEU	REF-INT-NEU	REF-SEN-NEU
ACT-INT-NEU	ACT-SEN-NEU	***GLO-INT-NEU***
GLO-SEN-NEU	SEQ-INT-NEU	SEQ-SEN-NEU
GLO-ACT-VIS-NEU	GLO-ACT-VER-NEU	GLO-REF-VIS-NEU
GLO-REF-VER-NEU	SEQ-ACT-VIS-NEU	SEQ-ACT-VER-NEU
SEQ-REF-VIS-NEU	SEQ-REF-VER-NEU	SEN-ACT-VIS-NEU
SEN-ACT-VER-NEU	***SEN-REF-VIS-NEU***	SEN-REF-VER-NEU
INT-ACT-VIS-NEU	INT-ACT-VER-NEU	INT-REF-VIS-NEU
INT-REF-VER-NEU	INT-GLO-VIS-NEU	INT-GLO-VER-NEU
INT-SEQ-VIS-NEU	INT-SEQ-VER-NEU	SEN-GLO-VIS-NEU
SEN-GLO-VER-NEU	SEN-SEQ-VIS-NEU	SEN-SEQ-VER-NEU
GLO-ACT-INT-NEU	GLO-ACT-SEN-NEU	GLO-REF-INT-NEU
GLO-REF-SEN-NEU	SEQ-ACT-INT-NEU	SEQ-ACT-SEN-NEU
SEQ-REF-INT-NEU	SEQ-REF-SEN-NEU	

6.4. Decision Unit

The output of the decision unit is the learner model. The learner model is presented with four dimensions, where each dimension represent learner's most dominant characteristic for that dimension. The learner model is obtained using decision unit as follows: After classifying all the learner profile table rows, a set of possible labels for each row as stated in the previous section, is obtained. Among them, the label with the highest confidence is assigned to the respective row as shown by an example in Table 6.7. Then, the total score for each pole is calculated using class label counts.

Table 6.7 Calculating total scores for each strong side and balanced case of each dimension

DIMENSION NAME OF LEARNING STYLE MODEL	TOTAL SCORE (# DIMENSIONNAME)
GLOBAL (GLO) SEUENTIAL (SEQ)	#GLO = SUM OF COUNTS OF CLASS LABELS INCLUDING "GLO" #SEQ = SUM OF COUNTS OF CLASS LABELS INCLUDING "SEQ"
ACTIVE (ACT) REFLECTIVE (REF)	#ACT = SUM OF COUNTS OF CLASS LABELS INCLUDING "ACT" #REF = SUM OF COUNTS OF CLASS LABELS INCLUDING "REF"
INTUITIVE (INT) SENSING (SEN)	#INT= SUM OF COUNTS OF CLASS LABELS INCLUDING "INT" #SEN = SUM OF COUNTS OF CLASS LABELS INCLUDING "SEN"
VISUAL(VIS) VERBAL(VER)	#VIS = SUM OF COUNTS OF CLASS LABELS INCLUDING "VIS" #VER = SUM OF COUNTS OF CLASS LABELS INCLUDING "VER"
BALANCED (NEU)	#NEU = SUM OF COUNTS OF CLASS LABELS INCLUDING "NEU" FOR THE CORRESPONDING DIMENSION

The learner is modeled according to his learning style using these total scores. Accordingly, in general, we get the following total score table format for each dimension of learning style as illustrated on Table 6.8.

Table 6.8 The total score table format for each dimension

Learning Style Dimension	Total Score of One Strong Side of The Learning Style Dimension	Total Score of the Other Strong Side of The Learning Style Dimension	Total Score of Balanced Case of The Learning Style Dimension
i	$\#Case_{i1}$	$\#Case_{i2}$	$\#Case_{i3}$

Each learning style dimension is diagnosed separately using the rules over total scores of the corresponding dimension. The rules for diagnosing learning style are illustrated in Table 6.9 by notifying the predetermined threshold as ε and difference of total scores as Δ:

Table 6.9 Rule Table for Learning Style Diagnose

Rules for Diagnosing Learner	Diagnosed Learning Style Dimension
minTotalScore = total number of learner profile table rows	

For learning style dimension i and for each column j, check the following:
if($\#Case_{ij}$<= minTotalScore)
{
minTotalScore= $\#Case_{ij}$
eliminatedColumnID=j;
}	

Let minTotalScore = Z and the other remaining two $\#Case_{ij}$, #X and #Y, where j!= eliminatedColumnID	
$(\Delta = \|(\#X) - (\#Y)\|) > \varepsilon$ and $\#X > \#Y$	X
$(\Delta = \|(\#X) - (\#Y)\|) > \varepsilon$ and $\#Y > \#X$	Y
$(\Delta = \|(\#X) - (\#Y)\|) \leq \varepsilon$	Z

In Table 6.9, threshold $\varepsilon = \varepsilon_{predetermined\_reference} * ((\#X) + (\#Y))$, where

$$\varepsilon_{predetermined\_reference} = (|(\#X) - (\#Y)|)_{predetermined\_reference} / ((\#X) + (\#Y))_{predetermined\_reference}$$

Predetermined reference is one of the total score tables of learners, selected according to the minimum difference value. Minimum difference value is defined as the difference between total scores of either poles or a pole and balanced case of a dimension.

As stated before, we suggest using this learner model, in consecutive logons of the same learner to the system. Hence, we can use this learner model in filtering results of a submitted generic query, such that the learner gets query results, which are more related and appropriate to his learning characteristics. Such a filtering can be done application specific, since we have to know the interdependency between attributes and the learning style dimension.

6.5. Experimental Results

To evaluate learner model results obtained with the proposed approach; i.e. the correctness of identifying learning styles based on FSLSM by using our proposed automatic learner model, we selected ILS [73] results as reference, since ILS questionnaire is an often used and well investigated instrument to identify learning styles based on the FSLSM. [39] Summarizes studies concentrated on analyzing the response data of the ILS questionnaire regarding the distribution of preferences for each dimension and also on verifying the reliability and validity of the instrument. Although there are a few studies [82], [83], where open issues arose, like weak reliability and validity and also dependencies between some learning styles, [39] concluded that the ILS questionnaire is a reliable and valid instrument and suitable for identifying learning styles according to FSLSM.

ILS is a 44 item forced-choice questionnaire, developed for identifying learning styles based on the FSLSM. As explained before, each learner has a preference on a pole for each of the four dimensions. The strength of preferences on a pole for each dimension is described with a score value ranged between 1 and 11with steps +/- 2 [28], [57].This range arises from 11 forced- choice items associated for each dimension. Each answer can have either +1 (answer a) or -1 (answer b) value. Answer (a) corresponds to the preference for the first pole of each dimension (active, sensing, visual, or sequential) and answer (b) to the second pole of each dimension (reflective, intuitive, verbal, or global). A sample, filled ILS questionnaire and its result can be seen in Appendix-B.

Furthermore, as stated on the questionnaire results page on the web, scores are presented on only one pole of a dimension and score values have the following meanings: If the score values range between 1-3, then the learner is balanced on the two poles of this dimension. If the score values range between 5-7, then the learner has more tendencies towards this scored pole of the dimension. If the score values range between 9-11, then the learner has strong characteristics in this scored pole of the dimension.

In our experimental results, we used three possibilities for each dimension as strong poles of the dimension (corresponding to ILS range 5-11) and balanced; i.e. neutral (corresponding to ILS range 1-3). Accordingly, the four dimensions of learning style are evaluated as follows: perception dimension as (SEN/INT/NEU), input dimension as (VIS/VER/NEU), understanding dimension as (GLO/SEQ/NEU), and processing dimension as (ACT/REF/NEU).

Regarding the experiment of the proposed approach, the training data set is prepared from LSDOs collected from 10 graduate students. These 10 students are selected such that each of them has a different learning style in order to enrich the training data set. There are 244 labeled rows in training data set. The properties of training data set are illustrated in Table 6.3. There are common selections of different students causing the rows belong to multiple labels. Therefore, we convert these 244 labeled rows to 68 multi-labeled rows, where the labels are NEU, REF-NEU, SEN-

VIS-NEU, INT-VIS-NEU, SEQ-NEU, VIS-NEU, REF-SEQ-NEU, GLO-INT-NEU, GLO-ACT-VIS-NEU, and SEN-REF-VIS-NEU.

First, for each learner, we constructed learner profile from learner selected data objects using our conversion unit.

Then, for each learner in the training set, we labeled his/her learner profile rows, which will be training set instances, with a class label from table 4.14 corresponding his/her learning style known by ILS result. Using obtained performance results of training data set with different classification algorithms in conjunction with multi-label classification methods, as shown on Table 6.4 and Table 6.5, we observed that minimum hamming loss, maximum accuracy and precision build our selected classifier with the training data set.; i.e. NBTree classifier in conjunction with BR.

For the test part, a group of 30 graduate students corresponding to 599 unlabeled rows is used. For each learner in the test data set, we processed learner profile rows over the clustering unit to obtain decision unit input. Hence, we get prediction label set for these unlabeled rows using the classifier of clustering unit. Then we assigned each unlabeled row the predicted class label with the highest confidence value. So, the input of learner model is ready.

An example is as shown in Table 6.10 for a sample tested learner with 18 test instances.

Table 6.10 Output of Clustering Unit

Tested Learner's Instances	**Predicted Label Set**	**Assigned Predicted Label with the highest Confidence**
1	{REF-NEU, SEQ-NEU}	SEQ-NEU
2	{REF-NEU,SEQ-NEU, VIS-NEU, INT-VIS-NEU, REF-SEN-VIS-NEU }	VIS-NEU
3	{ REF-NEU }	REF-NEU
4	{NEU,REF-NEU,VIS-NEU }	NEU
5	{ NEU,REF-NEU}	NEU
6	{REF-NEU,SEQ-NEU,VIS-NEU, REF-SEN-VIS-NEU }	REF-NEU
7	{REF-NEU,VIS-NEU,INT-VIS-NEU }	REF-NEU
8	{ NEU,REF-NEU,SEQ-NEU,VIS-NEU, INT-VIS-NEU, GLO-ACT-VIS-NEU,REF-SEN-VIS-NEU }	SEQ-NEU
9	{ NEU,REF-NEU,SEQ-NEU,VIS-NEU, INT-VIS-NEU, GLO-ACT-VIS-NEU,REF-SEN-VIS-NEU }	SEQ-NEU
10*	{ NEU , VIS-NEU}	NEU
11*	{ NEU , VIS-NEU , INT-VIS-NEU}	NEU
12	{ VIS-NEU }	VIS-NEU
13	{ VIS-NEU, GLO-INT-NEU}	VIS-NEU
14	{ SEQ-NEU, VIS-NEU }	VIS-NEU
15	{ SEQ-NEU, VIS-NEU }	VIS-NEU
16	{ SEQ-NEU, VIS-NEU, INT-VIS-NEU }	VIS-NEU
17	{ NEU , VIS-NEU}	VIS-NEU
18	{ REF-NEU , VIS-NEU , INT-VIS-NEU}	REF-NEU

The instances marked with * are having equal-confident predicted class labels. In such cases, we assigned the first label having most distinguishable characteristics among the highest equal-confident class labels. For example if NEU and VIS-NEU are equal-confident predictions, then we select VIS-NEU, which is a more distinguishable label with respect to label NEU.

Then using our learner model, we diagnosed learning style of the learner. The diagnosing learning style process is shown for this sample learner is illustrated in Table 6.11.

Table 6.11 Diagnosing learning style process inside learner model using predefined rules for diagnosing learning style with $\varepsilon_{predetermined\_reference} = 0.23$

Perception Dimension (SEN/INT/NEU)	**Total # of SEN**	**Total # of INT**	**Total # of NEU**	**DIAGNOSE**
	0	0	18	NEU
Input Dimension (VIS/VER/NEU)	**Total # of VIS**	**Total # of VER**	**Total # of NEU**	**DIAGNOSE**
	7	0	11	NEU
Understanding Dimension (GLO/SEQ/NEU)	**Total # of GLO**	**Total # of SEQ**	**Total # of NEU**	**DIAGNOSE**
	0	3	15	NEU
Processing Dimension (ACT/REF/NEU)	**Total # of ACT**	**Total # of REF**	**Total # of NEU**	**DIAGNOSE**
	0	4	14	NEU

In Table 6.12 and Table 6.13, we present the comparison the ILS questionnaire result and diagnosed learning style via our learner model for each tested learner. The bold italic parts on Table 6.12 and Table 6.13 illustrate mismatches.

Table 6.12 Comparison of ILS results to Learner Model Results -1

Tested Learner	No. of Test Instances	Perception (SEN/INT/NEU)		Input (VIS/VER/NEU)	
		ILS result	Learner Model result	ILS result	Learner Model result
1	36	NEU	NEU	***NEU***	***VIS***
2	34	NEU	NEU	***NEU***	***VIS***
3	36	NEU	NEU	***NEU***	***VIS***
4	47	NEU	NEU	VIS	VIS
5	37	***INT***	***NEU***	VIS	VIS
6	20	***SEN***	***NEU***	VIS	VIS
7	14	NEU	NEU	VIS	VIS
8	6	NEU	NEU	***NEU***	***VIS***
9	6	***SEN***	***NEU***	VIS	VIS
10	8	INT	NEU	***NEU***	***VIS***
11	18	NEU	NEU	NEU	NEU
12	9	NEU	NEU	***NEU***	***VIS***
13	12	NEU	NEU	VIS	VIS
14	17	SEN	NEU	VIS	VIS
15	6	NEU	NEU	VIS	VIS
16	20	NEU	NEU	***NEU***	***VIS***
17	15	NEU	NEU	VIS	VIS
18	22	***SEN***	***NEU***	***NEU***	***VIS***
19	26	NEU	NEU	NEU	NEU
20	19	NEU	NEU	***NEU***	***VIS***
21	10	NEU	NEU	VIS	VIS
22	8	NEU	NEU	VIS	VIS
23	28	***INT***	***NEU***	***VER***	***NEU***
24	32	NEU	NEU	***NEU***	***VIS***
25	10	NEU	NEU	***NEU***	***VIS***
26	15	NEU	NEU	***NEU***	***VIS***
27	22	***INT***	***NEU***	VIS	VIS
28	18	***SEN***	***NEU***	VIS	VIS
29	24	NEU	NEU	***NEU***	***VIS***
30	16	NEU	NEU	NEU	NEU

Table 6.13 Comparison of ILS results to Learner Model Results -2

Tested Learner	No. of Test Instances	Understanding (GLO/SEQ/NEU)		Processing (ACT/REF/NEU)	
		ILS result	Learner Model result	ILS result	Learner Model result
1	36	NEU	NEU	NEU	NEU
2	34	NEU	NEU	NEU	NEU
3	36	NEU	NEU	NEU	NEU
4	47	NEU	NEU	NEU	NEU
5	37	NEU	NEU	NEU	NEU
6	20	NEU	NEU	NEU	NEU
7	14	***GLO***	***NEU***	***GLO***	***NEU***
8	6	***SEQ***	***NEU***	***SEQ***	***NEU***
9	6	NEU	NEU	NEU	NEU
10	8	***GLO***	***NEU***	***GLO***	***NEU***
11	18	NEU	NEU	NEU	NEU
12	9	***SEQ***	***NEU***	***SEQ***	***NEU***
13	12	NEU	NEU	NEU	NEU
14	17	NEU	NEU	NEU	NEU
15	6	NEU	NEU	NEU	NEU
16	20	***SEQ***	***NEU***	***SEQ***	***NEU***
17	15	NEU	NEU	NEU	NEU
18	22	NEU	NEU	NEU	NEU
19	26	NEU	NEU	NEU	NEU
20	19	NEU	NEU	NEU	NEU
21	10	***GLO***	***NEU***	***GLO***	***NEU***
22	8	NEU	NEU	NEU	NEU
23	28	NEU	NEU	NEU	NEU
24	32	NEU	NEU	NEU	NEU
25	10	***SEQ***	***NEU***	***SEQ***	***NEU***
26	15	NEU	NEU	NEU	NEU
27	22	NEU	NEU	NEU	NEU
28	18	***GLO***	***NEU***	***GLO***	***NEU***
29	24	NEU	NEU	NEU	NEU
30	16	***SEQ***	***NEU***	***SEQ***	***NEU***

Using the information on Table 6.12 and Table 6.13, the following statistics can be made as presented on Figure 6.4 and Figure 6.5.

Figure 6.4 shows the match ratio by presenting the relation between the number of dimensions of the learning style model, for which ILS and our learner model results are the same and the number of learners. Looking at Figure 4.6., it can be concluded that, in the worst case for 3 learners out of 30, only one dimension of the learning style is found as the same as ILS results. On the other hand, for 5 learners out of 30, all dimensions of the learning style are found as the same as ILS results.

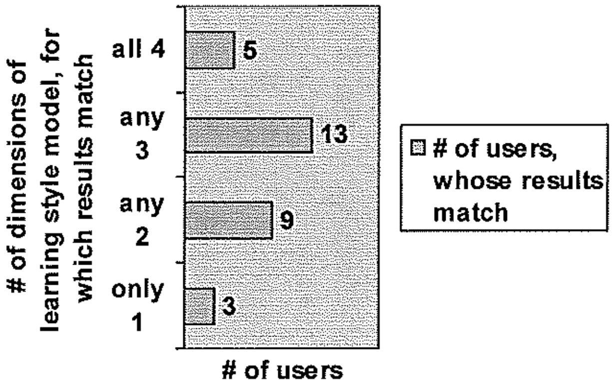
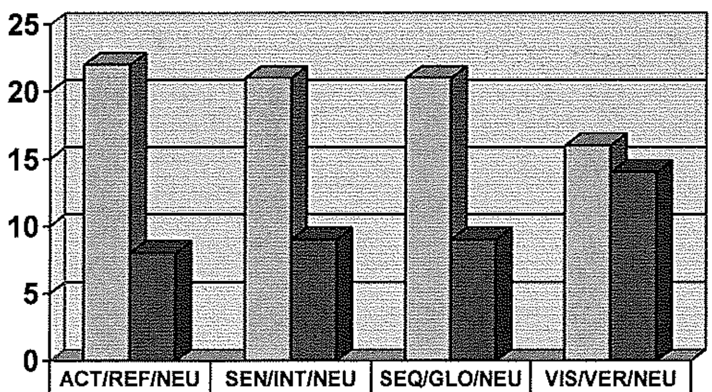


Figure 6.4 Match ratio of ILS results and our proposed learner model results

Figure 6.5 shows the number of misses and matches for each dimension of learning style model. Looking at Figure 6.5, it can be concluded that for these experimental learner group, the most misses occur in input dimension; i.e. the success ratio is 53,3%. On the other hand, processing dimension has the success ratio of 70% and the remaining two dimensions, namely understanding and perception dimensions have the same hit success ratio of 73,3%.



	ACT/REF/NEU	SEN/INT/NEU	SEQ/GLO/NEU	VIS/VER/NEU
# of Matches	22	21	21	16
# of Misses	8	9	9	14

Figure 6.5 Miss ratio of dimensions of learning style model

6.6. Critique of Experimental Results

In order to get an idea about these success ratios for this experimental group, we compare our FSLSM learning style diagnose results with the results of other works available in the literature, the conclusion is as presented in Table 6.14.

Table 6.14 Comparison Results

	Perception Dimension (SEN/INT)	Understanding Dimension (GLO/SEQ)	Processing Dimension (ACT/REF)	Input Dimension (VIS/VER)
Our Proposed Work (30 tested learner)	73.3%	73.3%	70%	53.3%
[51](27 tested learner)	77%	63%	58%	Considered as not suitable to evaluate in this work
[46](25 test data sets)	77%	63.4%	56.7%	84%

Hence, based on the experimental results, we can conclude that the proposed work is more effective in understanding and processing dimensions than the other works. However, although the succes ratio is fairly close to the other works in perception dimension, input dimension is overwhelmingly low. Considering that, one of the available works even not evaluated this dimension and the other work uses directly related behavioural patterns for this dimension, the low success ratio in this dimension can be categorized as tolerable.

Furthermore, as mentioned before, the other works investigate learner's behavioural patterns using different techniques like decision trees or Bayesian networks. The difference of the proposed work is to diagnose learning style of learners only using learner interested data objects using NBTree in conjunction with Binary Relevance (BR) classifier, a multilabel classification method. As presented in Table 6.4 and Table 6.5, the proposed classification method, provides the minimum hamming loss, maximum accuracy and precision, where hamming loss is a performance evaluation measure, representing prediction errors(an incorrect label is predicted) and missing errors(a label is not predicted). Therefore, when a classification method has a good performance in prediction, then it has to have a low value for hamming loss and a high value for accuracy and precision.

On the other hand, considering Table 6.4 and Table 6.5, multi-label classification method LP is most successful with ID3 classifier. However, in this case, the confidence levels of the predicted test set instances varies as 1 or 0, which are not helpful in assigning a predicted label for the test set instance. Whereas, the confidence levels vary between 0.0 and 9.99 in the proposed approach. Similarly, multi-label classification method named method-6 is most successful with C4.5 classifier. In this case, the confidence levels of the predicted test set instances varies between 0.0 and 1.0, where multiple labels get 1.0 value and this is also not helpful in assigning a predicted label for the test set instance.

Furthermore, most of the trained and tested learners have NEU characteristic in processing, perception and understanding dimensions and VIS characteristic in input dimension of FSLSM in ILS results as presented in Table 6.12 and Table 6.13. Additionally, the classification with the worst performance occurs for NB classifier in conjunction with method-6, where hamming loss is maximum and accuracy, precision and recall values are low as presented on Table 6.4 and Table 6.5. All of the tested learner instances are predicted as VIS-NEU with this classification method. However, due to the dominant VIS characteristic in input dimension and NEU characteristic in other dimensions of FSLSM of the tested learners, the low performance of this classification method; i.e. NB classifier in conjunction with method-6; in prediction is not obviously observable. To observe the low performance of this classification method in prediction more explicitly, let us consider the following trained and tested learner characteristics as presented on Table 6.15 and Table 6.16.

Table 6.15 Another Comparison of ILS Results to Learner Model Results-1

Tested Learner	No. of Test Instances	Perception (SEN/INT/NEU)		Input (VIS/VER/NEU)	
		ILS result	Learner Model result	ILS result	Learner Model result
1	36	***NEU***	***SEN***	VIS	VIS
2	34	NEU	NEU	***NEU***	***VIS***
3	36	NEU	NEU	***NEU***	***VIS***
4	47	SEN	SEN	VIS	VIS
5	37	SEN	SEN	VIS	VIS
6	20	***INT***	***NEU***	VIS	VIS
7	14	NEU	NEU	***NEU***	***VIS***
8	6	***INT***	***NEU***	***NEU***	***VIS***
9	6	NEU	NEU	VIS	VIS
10	8	SEN	SEN	VIS	VIS

Table 6.16 Another Comparison of ILS Results to Learner Model Results-2

Tested Learner	No. of Test Instances	Understanding (GLO/SEQ/NEU)		Processing (ACT/REF/NEU)	
		ILS result	Learner Model result	ILS result	Learner Model result
1	36	***GLO***	***NEU***	***ACT***	***NEU***
2	34	SEQ	SEQ	REF	REF
3	36	SEQ	SEQ	NEU	NEU
4	47	NEU	NEU	NEU	NEU
5	37	NEU	NEU	NEU	NEU
6	20	NEU	NEU	NEU	NEU
7	14	SEQ	SEQ	REF	REF
8	6	***GLO***	***NEU***	***NEU***	***REF***
9	6	***GLO***	***NEU***	***ACT***	***NEU***
10	8	NEU	NEU	REF	REF

In this case, the success ratio in diagnosing learning style with the learner model using NBTree classification with BR classifier become as follows: 70% in perception, understanding and processing dimensions and 60% in input dimension.

Furthermore, in this case, all of the tested learners are predicted as SEN-VIS-NEU by the classification method with the worst performance; i.e. NB classifier in conjunction with method-6. Hence, the success ratio in diagnosing learning style with the learner model using this classification method become as follows: 30% in perception dimension, 60% in input dimension, 40 % in understanding dimension and 50% in processing dimension.

Hence, we can conclude that, the proposed learner model classification method (NBTree classifier in conjunction with BR classifier) has a moderate 70-72% success ratio for perception, understanding and processing dimensions, whereas 55-56% success ratio for input dimension, which is not very effective. Considering that the proposed technique takes into account common selections for different learning style characteristics and handles this problem without observing learner behaviour in time scale differently from the available works, it is worth to expand this work from the cognitive point of view in the future.

CHAPTER 7

AN APPLICATION SCENARIO FOR THE OVERALL SYSTEM

To illustrate the working principle of the proposed e-learning system, an online course is considered as a sample application scenario.

In this sample scenario, as local resources, relational database tables are used. The attributes "Learning ResourceType" and "RequiredDegreeOfCollaboration" are IEEE LOM metadata fields and the other two attributes are added as a result of our requirements. The entries for these attribute values are the corresponding metadata values. An example for the structure of the local data resource is presented in Table 7.1.

Table 7.1 Structure of the local data resource

LOID	**LEARNING RESOURCE TYPE**	**REQUIREDDEGREEOFCOLLABORATION**	**TYPEOFCOVERAGE**	**THEARTOFPRESENTINGCONTENT**
1	EXERCISE	INDIVIDUAL WORK	PRACTICAL-KNOWN METHODS, APPLICABLE TO REAL WORLD	SEQUENTIALLY-STEP BY STEP

Further, we assume that we have two different local data resources for this application scenario, such that one is prepared in English and the other is prepared in Turkish.

The scenario starts, when both Turkish students A and B submit a query about the on-line course "cell" to the system. Since the proposed application layer is OGSA-DQP GUI, the query is submitted in the form of Object Query Language (OQL) syntax as shown on Figure 7.1.

```
"%print select cell.* from cell in ENG;"
```

Figure 7.1 Submitted Query

After query submission, for personalization part, we send the query to OGSA-DQP.

The returned result set will be presented from OGSA-DQP GUI to both learners A and B as shown in Table 7.2.

Table 7.2 Returned Result Set

LOID	LEARNING RESOURCE TYPE	REQUIREDDEGREEOF COLLABORATION	TYPEOFCOVERAGE	THEARTOFPRES ENTINGCONTENT
1	VISUAL RESOURCE	TEAM WORK	PRACTICAL-KNOWN METHODS, APPLICABLE TO REAL WORLD	GLOBALLY-WITH THE BIG PICTURE OF SUBJECT
2	NARRATIV E TEXT	ALONE	THEORETICAL	SEQUENTIALLY, STEP BY STEP
3	NARRATIV E TEXT	TEAM WORK	THEORETICAL	GLOBALLY-WITH THE BIG PICTURE OF SUBJECT
4	VISUAL RESOURCE	TEAM WORK	PRACTICAL-KNOWN METHODS, APPLICABLE TO REAL WORLD	SEQUENTIALLY, STEP BY STEP

Then, each learner selects the rows, on which he/she is mostly interested in. So, for each learner, we got learner interested data objects as input for the proposed automatic learner modelling module. In this scenario, due to the structure of the local data resources, we can bypass keyword finder and mapper stages of the proposed conversion unit.

In this scenario, the selected data objects directly correspond to learner profile as shown in Table 7.3 for learner A and in Table 7.4 for learner B.

Table 7.3 Learner Profile for Learner A

LEARNING RESOURCE TYPE	REQUIREDDEGREEOF COLLABORATION	TYPEOFCOVERAGE	THEARTOF PRESENTING CONTENT
VISUAL RESOURCE	TEAM WORK	PRACTICAL-KNOWN METHODS, APPLICABLE TO REAL WORLD	GLOBALLY-WITH THE BIG PICTURE OF SUBJECT
VISUAL RESOURCE	TEAM WORK	PRACTICAL-KNOWN METHODS, APPLICABLE TO REAL WORLD	SEQUENTIALLY,STEP BY STEP

Table 7.4 Learner Profile for Learner B

LEARNING RESOURCE TYPE	REQUIREDDEGREEOF COLLABORATION	TYPEOFCOVERAGE	THEARTOF PRESENTING CONTENT
NARRATIVE TEXT	ALONE	THEORETICAL	SEQUENTIALLY,STEP BY STEP
NARRATIVE TEXT	TEAM WORK	THEORETICAL	GLOBALLY-WITH THE BIG PICTURE OF SUBJECT

In the next step, for clustering unit, we assume that we have built our NBTree classifier in conjunction with Binary Relevance (BR) classifier, with previously prepared training set as an offline preprocess. Then, for each learner, we classify learner profile rows; i.e. our test instances using our clustering unit and obtain the predicted label set. To determine learner model input, we assign each test instance a predicted label with the highest confidence. The predicted label set and the assigned label are illustrated on Table 7.5 for learner A and on Table 7.6 for learner B.

Table 7.5 Output of Clustering Unit for Learner A

Instances of Tested LEARNER A	**Predicted Label Set**	**Assigned Predicted Label with the highest Confidence**
1	{NEU,SEQ-NEU,VIS-NEU,REF-SEN-VIS-NEU}	VIS-NEU
2	{NEU, VIS-NEU}	VIS-NEU

Table 7.6 Output of Clustering Unit for Learner B

Instances of Tested LEARNER B	**Predicted Label Set**	**Assigned Predicted Label with the highest Confidence**
1	{NEU,VER-NEU}	VER-NEU
2	{NEU,INT-SEQ-NEU,VER-NEU}	VER-NEU

Hence, we get the total scores for each side of each dimension of the learning style of the learner. Using these total score values, we can model learner, according to his dominant side in each dimension of the learning style. Total score value distribution is presented on Table 7.7 for learner A and on Table 7.8 for learner B.

Table 7.7 Total scores on each side of each dimension for learner A

#GLO	0	**#SEQ**	0	**#NEU**	2
#SEN	0	**#INT**	0	**#NEU**	2
#ACT	0	**#REF**	0	**#NEU**	2
#VIS	2	**#VER**	0	**#NEU**	0

Table 7.8 Total scores on each side of each dimension for learner B

#GLO	0	**#SEQ**	0	**#NEU**	2
#SEN	0	**#INT**	0	**#NEU**	2
#ACT	0	**#REF**	0	**#NEU**	2
#VIS	0	**#VER**	2	**#NEU**	0

According to Table 7.7, we can conclude that learner A is dominantly visual, balanced on the remaining dimensions. Similarly, according to Table 7.8, we can conclude that learner B is dominantly verbal, balanced on the remaining dimensions.

So far, learning styles of learner A and learner B are determined. Now, for the submitted queries of these learners, we return the necessary results for these learners. For example, considering only that the learner A dominantly visual, when the learner A again submits cell as input query, the query is automatically reformulated to the following one illustrated on Figure 7.2

```
"%print select cell.*  from cell in ENG where cell.LearningResourceType=visual resource;"
```

Figure 7.2 Reformulated Query before Data Integration Module for learner A

Similarly, considering only that the learner B dominantly verbal, when the learner B again submits cell as input query, the query is automatically reformulated to the following one illustrated on Figure 7.3

```
"%print select cell.* from cell in ENG where cell.LearningResourceType=verbal resource;"
```

Figure 7.3 Reformulated Query before Data Integration Module for learner B

Then, data integration module starts to look for schema matching and reformulate the query according to the found matches. Hence, the reformulated query to be send to OGSA-DQP is as given in Figure 7.4 for learner A and in Figure 7.5 for learner B.

```
"%print select cell.*, hucre.* from cell in ENG, hucre in TR where
cell.LearningResourceType=visual resource and
hucre.OgretimKaynagiTipi= gorsel kaynak;"
```

Figure 7.4 Reformulated Query after Data Integration Module for learner A

```
"%print select cell.*, hucre.* from cell in ENG, hucre in TR where
cell.LearningResourceType=narrative text and
hucre.OgretimKaynagiTipi= yazılı metin;"
```

Figure 7.5 Reformulated Query after Data Integration Module for learner B

The returned result appearing on OGSA-DQP GUI for learner A will be as presented in Table 7.9:

Table 7.9 Returned Result Set After Personalization and Data Integration for learner A

LOID	LEARNING RESOURCETYPE	REQUIREDDEGREE OFCOLLABORATION	TYPEOFCOVERAGE	THEARTOF PRESENTING CONTENT
1	VISUAL RESOURCE	TEAM WORK	PRACTICAL-KNOWN METHODS, APPLICABLE TO REAL WORLD	GLOBALLY-WITH THE BIG PICTURE OF SUBJECT
4	VISUAL RESOURCE	TEAM WORK	PRACTICAL-KNOWN METHODS, APPLICABLE TO REAL WORLD	SEQUENTIALLY,S TEP BY STEP
LOID	**BILGI KAYNAGITIPI**	**ÇALIŞMATIPI**	**İÇERIKTIPI**	**İÇERIKSUNMAŞ EKLI**
11	GORSEL KAYNAK	GRUP ÇALIŞMASI	PRATIK	TÜMDENGELIM
12	GORSEL KAYNAK	BIREYSEL ÇALIŞMA	PRATIK	TÜMEVARIM

The returned result appearing on OGSA-DQP GUI for learner B will be as presented in Table 7.10:

Table 7.10 Returned Result Set After Personalization and Data Integration for learner B

LOID	LEARNING RESOURCETYPE	REQUIREDDEGREEOF COLLABORATION	TYPEOFCOVERAGE	THEARTOF PRESENTING CONTENT
1	NARRATIVE TEXT	ALONE	THEORETICAL	SEQUENTIALLY,STEP BY STEP
4	NARRATIVE TEXT	TEAM WORK	THEORETICAL	GLOBALLY-WITH THE BIG PICTURE OF SUBJECT
LOID	**BILGIKAYNAGI TIPI**	**ÇALIŞMATIPI**	**İÇERIKTIPI**	**İÇERIKSUNMA ŞEKLI**
11	YAZILI METIN	BIREYSEL ÇALIŞMA	TEORIK	TÜMEVARIM
12	YAZILI METIN	GRUP ÇALIŞMASI	TEORIK	TÜMDENGELIM

So far, we illustrated the workflow of the proposed e-learning system based on the OGSA infrastructure model.

7.1. Application To Web

Personalization in using online search engine for educational purposes is also investigated in the literature [101]-[104]. Personalized Search Tool for Teachers ([101] – [104]) called PoSTech is used as front-end of search engines that allows teachers to specify their search according to their own pedagogical vocabulary. This tool is designed for teachers as a mediator between educational information provided by search engines and students having different requirements. The current version of this tool uses AltaVista as search engine and mediation process is done using a filtering mechanism based on adding keywords to the original query.

Regarding the easy integrability of the obtained, proposed learner model for Web-based search, the following example is considered. In this example, the assumed learner is diagnosed as a verbal learner using the proposed learner model as described above. Then, for filtering Web-based query results of this verbal learner, we integrated the diagnosed learner model via a front-end tool to search engine using a simple filtering mechanism. In the designed front-end tool, we hold the diagnosed learning style information for the logged user as verbal or visual. Accordingly, we filter the original query before submitting to the search engine using some keywords. In the given example, we used Google as search engine and hence keyword used for filtering are as follows: For the verbal learner, we add the keyword "+filetype:pdf" to the original query. For the visual learner, we add the keyword "+picture" to the original query.

Hence, we provide the verbal learner, query results containing dominantly textual information rather than query results, which also contain visual material. Figure 5.6 illustrates the comparison of "cloning" search result over "Google" search engine for a verbal learner, with and without integration of the diagnosed learner model to the front-end tool.

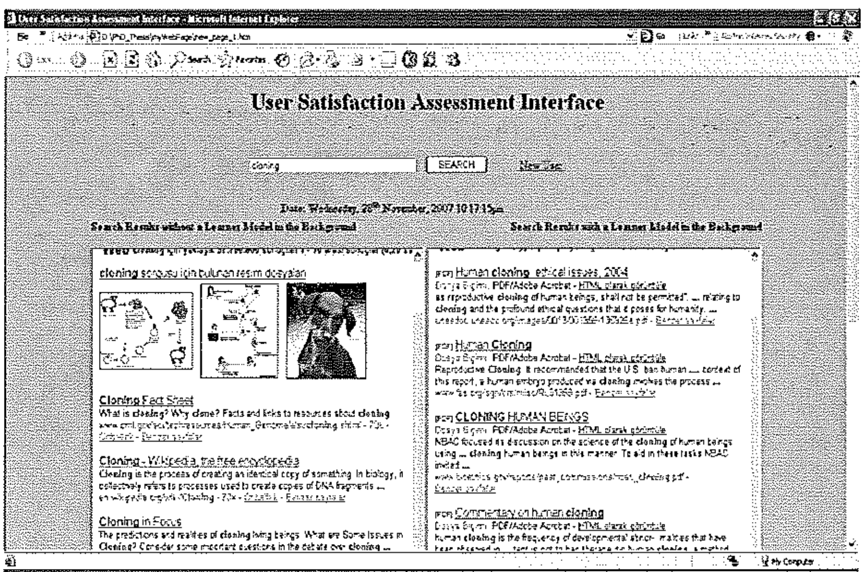


Figure 7.6 Comparison of search results with and without a learner model

On the left hand side of figure 7.6, the ordinary search results, obtained without a learner model, are presented. As illustrated on the right hand side of the figure 5.6, for the verbal learner, search results, obtained with the learner model, are presented. In this case, we filter material including visual material dominantly and provide the verbal learner the material containing textual information dominantly.

Regarding the learner satisfaction, although learner satisfaction assessment for learner models is investigated as another research area in the literature in the educational and social sciences domain, to give an idea we observed simply the following result in the given example: In the case, without using the learner model, 11 results are returned on the first page by the search engine and none of them are .pdf file. Whereas, in the case, using the learner model, 10 results are returned on the first page by the search engine and 9 of them are .pdf file.

Furthermore, the other dimensions of learning style like global/sequential, sensing/intuitive or active/reflective is not included in this example. Since, filtering query results according to these dimensions require a further detailed pedagogical and cognitive study.

CHAPTER 8

CONCLUSION AND FUTURE WORK

We suggested a complete e-learning framework composed of mainly two parts.

In the first part, we rebuild an infrastructure model based on grid technology for an e-learning system, namely OGSA-DAI and OGSA-DQP reference implementations. Using such kind of an infrastructure model based on grid technology in an e-learning framework brings the advantage of the availability of contextualization information for all tools. However, when using traditional LMSs, context is derived from the LMS's knowledgebase and applicable only within the LMS. To overcome this limit of LMSs,LMS can be transformed to grid technology. Hence, we introduced the gridtechnology usage in an infrastructure model. Our contribution to this open source infrastructure model reference implementation is a data integration module, which overcome language heterogeneity between local data resource structure.

Then, in the second part, for one of the most important features, namely personalization in e-learning systems, we addressed the personalization issue. We proposed automatic learner modeling based on the learner's learning style for e-learning. Automatic learner modeling is a hot subject in the literature and there are various automatic learner models designed with different techniques available in the literature. However, these learner models are dependent on underlying infrastructure and developed using the log information of co working LMSs or specially developed user interface programs. In the proposed approach, we addressed the following issue: There is a lack in designing an automatic learner model to diagnose learner's learning style using only the content of the data objects selected by the learner but not learner's other behaviour observed in time scale. Such kind of an automatic learner model design will bring the independency from a background, underlying LMS or special user interfaces. As a result, simplicity and easy integrability of such a learner model to an infrastructure model or LMS or just Web based search can be obtained. So, in the proposed approach, initially the learner interest is collected explicitly using generic queries. Then, the proposed learner profile is constructed using the proposed conversion unit based on keyword-mapping. From the learner profile the learner model is built using the proposed clustering unit. This clustering unit uses NBTree classification in conjunction with Binary relevance Classifier based on the Felder-Silverman learning style model.

On the other hand, the performance of such an automated learner model is unknown and should be observed. With the proposed approach, we also addressed this issue. The experimental results show the match ratio between the obtained learner's learning style using the proposed learner model and those obtained by the questionnaires traditionally used for learning style assessment. Observing the experimental results of the proposed automatic learner model, its performance is as follows: For input dimension, performance is 53.3%. On the other hand, processing dimension has the success ratio of 70% and the remaining two dimensions, namely understanding and perception dimensions have the same hit success ratio of 73.3%.

We contributed to literature from a different point of view regarding automatic learner modeling technique. Differently from the available works in the literature, our design uses sometimes not totally individually independent instead correlated input to diagnose different dimensions of the learning style. Furthermore, our design also technically permits to multi-label classification cases. These cases occur, since learners with different learning style can have common data object selections. Our design handles such kind of cases by evaluating different available multi-label methods with different suitable classification algorithms in our design. So, we illustrated the performance of such a learner model covering multi-label classification and input interdependency.

We addressed the automatic learner modeling for learning style diagnosing using a different technique from the available works in the literature. We proposed a technique, which isolates us from the learner behaviour observed in time scale and concerns learner selected data objects. Furthermore, the input pattern for the proposed automatic learner model is correlated, whereas in the available works independent input patterns are used for diagnosing learning styles.

Furthermore, assessment of using automatic learner models is another research subject in the literature. As a further study, with the cooperation of educational science coworkers the assessment of the proposed automatic learner model can be realized over a controlled experimental group.

Also, unsupervised learning and pattern recognition technologies may be utilized to the benefit of clustering unit of the proposed automatic learner model.

REFERENCES

[1] "e-Learning Interoperability Standards", Sun White paper, retrieved from http://www.sun.com/solutions/documents/whitepapers/ed_interop_standards.pdf, last visited on January 2002

[2] IEEE.Draft Standard For Learning Object Metadata (IEEE P1484.12.1) , 2002

[3] ADL. Sharable content object reference model (SCORM) 2004.Technical report, Advanced Distributed Learning, 2004

[4] "Trends and Issues in E-learning Infrastructure Development", A White Paper for alt-i-lab 2004, Prepared on behalf of DEST (Australia) and JISC-CETIS (UK), Principal Contributors: Kerry Blinco Jon Mason Neil McLean Scott Wilson, Version 2, 2004

[5] Foster, I., "What is the Grid? A Three Point Checklist", 2002

[6] Schopf, J.M., Nitzberg, B., " Grids: The Top Ten Questions ", Scientific Programming, special issue on Grid Computing,vol.10,no.2,pp.103-111, 2002

[7] Humphrey, M., Wasson, G.,Kiryakov, Y., Park,S., Del Vecchio, D. and Beekwilder, N., "Alternative Software Stacks for OGSA-based Grids", Proceedings of Supercomputing 2005, Seattle, WA, Nov 12-18, 2005

[8] "Simple Query Interface (SQI) for Learning Repositories", Specification, Version 0.7, Public Draft

[9] Kraan, W., "The cockroach of repository interoperability: Simple Query Interface", CETIS staff, 2004

[10] Simon, B., Massart, D., Assche, F. van, Ternier, S., Duval, E., Brantner, S. Olmedilla, D., Miklos, Z., "A Simple Query Interface for Interoperable Learning Repositories", WWW2005,Chiba,Japan, 2005

[11] "OGSA-DQP GUI", retrieved from http://www.ogsadai.org.uk/downloads/archive/ogsa-dqp_3.0, last visited on October 2005

[12] Ziegler, P., Dittrich, K.R., "Three Decades of Data Integration – All Problems Solved?", First International IFIP Conference on Semantics of a Networked World (ICSNW 2004)

[13] Miller, R.J. , Hernandez, M.A., Haas, L.M., Yan, L.-L., Ho, C.T.H., Fagin, R. and Popa, L., "The Clio Project: Managing Heterogeneity.", SIGMOD Record, vol.30,no.1, pp.78-- 83, 2001

[14] Do, H. ,Rahm, E., "COMA – A system for flexible combination of schema matching approaches", Proceedings of the 28 th VLDB Conference, Hong Kong, China, 2002

[15] "Automed: AutoMatic Generation of Mediator Tools for Heterogeneous Data Integration", retrieved from http://www.lkl.ac.uk/cms/index.php?option=com_content&task=view&id=129&Itemid=91, last visited on September 2005

[16] "Data Integration Using AutoMed, OGSA-DAI & DQP", retrieved from http://www.dcs.bbk.ac.uk/~lucas/talks/ISPIDER-2005-07-BBK.pdf, last visited on September 2005

[17] Zamboulis, L., "The AutoMed Query Processor", AutoMed Project Meeting, Imperial College, London, 2005

[18] Comito, C., Talia, D., Gounaris, A., Sakellariou, R., "Data integration and query reformulation in service-based Grids:Architecture and Roadmap", 2005

[19] Comito, C., Talia, D., "GDIS: A Service-Based Architecture for Data Integration on Grids", OTM Workshops 2004, LNCS 3292, pp. 88–98, Springer-Verlag Berlin Heidelberg, 2004

[20] Woehrer,A. and Brezany,P. and Min Tjoa, A., "Novel mediator architectures for Grid information systems", Future Generation Computer Systems, vol.21,no.1,pp.107-114, 2005

[21] Pankratius,V. and Vossen,G., "Towards E-Learning Grids: using Grid Computing in Electronic Learning", Proceeding of IEEE Workshop on Knowledge Grid and Grid Intelligence, Nova Scotia, Canada, pp.4-15, 2003

[22] Li, K.,Tsai,K.,Tsai,Y. and Wang,H., " Toward Design of E-Learning Platform in Grid Environments", Proceedings of the 2006 International Conference on Grid Computing & Applications, GCA 2006, pp. 213-220, Las Vegas, Nevada, USA, 2006

[23] Navarro, L.I., Such, M.M., Diaz, A.S., Martin, D.M. and Peco, P.P. , "Learning Units Design based in Grid Computing.", Fifth IEEE International Conference on Advanced Learning Technologies, pp. 562 – 566, 2005

[24] Marton, F. ," Phenomenography - research approach to investigating different understandings of reality.", Journal of thought, vol. 21, pp. 28-49, 1986

[25] Bedford, T. A.," Learning styles: a review of literature (first draft)", Toowoomba, OPACS, The University of Southern Queensland, 2004

[26] Coffield, F., Moseley, D., Hall, E., Ecclestone, K., "Learning styles and pedagogy in post-16 learning, A systematic and critical review", 2004

[27] Messick, S.,"The nature of cognitive styles: problems and promise in educational
practice." Educational Psychologist, vol.19, no.2, pp.59–74,1984

[28] Felder, R.M., Brent, R., "Understanding Student Differences", Journal of Engineering Education, vol. 94, no.1, pp.57-72 , 2005

[29] "Cpp - products: Myers-briggs type indicator (mbti) assessment", retrieved from http://www.cpp.com/products/mbti/, last visited on February 2007

[30] Felder, R. M., & Silverman, L. K., "Learning and teaching styles in engineering– education. Engineering Education", vol.78,no.7,p.674. (Including Author's Preface - June2002 by Richard M. Felder), 1988

[31] Kolb, D. A., Boyatzis, R. E., & Mainemelis, C., "Experiential learning theory: Previous research and new directions". In R. J. Sternberg & L. F. Zhang (Eds.), Perspectives on cognitive, learning, and thinking styles. NJ: Lawrence Erlbaum., 2000

[32] Herrmann, N., "The creative brain". Lake Lure, North Carolina, USA: The Ned Herrmann Group., 1989

[33] Dunn R., "The Dunn and Dunn learning style model and its theoretical Cornerstone.", In R Dunn and S Griggs (Eds.), Synthesis of the Dunn and Dunn learning styles model research: who, what, when, where and so what – the Dunn and Dunn learning styles model and its theoretical cornerstone, 1–6. New York: St John's University, 2003

[34] H.W. Boschloo, "Learning Styles Literature Review Virtual Assistant Project",2005

[35] Kolb, D. A., "Learning style inventory: self-scoring inventory and interpretation
booklet.", Boston, USA: McBer., 1985

[36] Herrmann International.," Herrmann international: The ned herrmann group", retrieved from http://www.hbdi.com, last visited on February 2007

[37] Herrmann-Nehdi, A.,"Training with the brain in mind: The application of brain dominance technology to teaching and learning." , In Training 2002 conference & expo. Atlanta, Georgia, USA., 2002

[38] Felder, R. M., "Reaching the second tier: Learning and teaching styles in college
science education.", College Science Teaching, vol.23,no.5, pp. 286-290,1993

[39] Felder, R. M., & Spurlin, J. ,"Applications, reliability and validity of the index of learning styles. ", Intl. Journal of Engineering Education, vol.21,no.1, pp.103-112, 2005

[40] Wang,L. and Bai,J.,"A synthesis analysis and comparison of text classification algorithms", retrieved from www.iro.umontreal.ca/~nie/IFT6255/Projets//Classification.ppt , last visited on March 2007

[41] Bennet,P.,"Introduction to Text Categorization ", retrieved from http://www.softlab.ntua.gr/facilities/public/AD/Text%20Categorization/Introduction%20to%20Text%20Categorization.ppt#256, last visited on March 2007

[42] "SVM", retrieved from http://nlp.stanford.edu/IR-book/ppt/lecture15-svm.pdf

[43] Mooney,R.J., "Text Categorization",retrieved from http://www.cs.utexas.edu/~mooney/ircourse/slides/TextCategorization.ppt, last visited on March 2007

[44] aDeNu Research Group (Santos, Olga C., Boticario, Jesus G., Rodriguez, A., Barrera,C.).,"Modeling Learner Interaction Preferences in dotLRN". OpenACS and .LRN Spring Conference, Vienna, 2007

[45] Carver, C.A., Howard, R.A. and Lane., W. D.,"Addressing different learning styles through course hypermedia.", IEEE Transactions on Education, vol.42,no. 1, pp. 33-38.,1999

[46] Cha, H.J., Kim, Y.S., Lee, J.H. and T. B. Yoon., "An Adaptive Learning System with Learning Style Diagnosis based on Interface Behaviors.",Workshop Proceedings of International Conference on E-Learning and Games, Hangzhou, China, 2006

[47] Denaux, R., Dimitrova, V., and L., Aroyo., " Interactive Ontology-Based Learner Modeling for Personalized Learning Content Management", SWEL'04 Workshop, 2004

[48] Dolog, P. and Nejdl,W., "Personalisation in Elena: How to cope with personalisation in distributed eLearning Networks.", In Proc. of International Conference on Worldwide Coherent Workforce, Satisfied Users - New Services For Scientific Information, Oldenburg, Germany , 2003

[49] dotLRN, retrieved from http://www.dotlrn.org, last visited on April 2008

[50] Dou, Z., Song, R. and Wen,J., "A Large-scale Evaluation and Analysis of Personalized Search Strategies.", Proceedings of the 16th international conference on World Wide Web, SESSION: Personalization pp.581-590, 2007

[51] García, P., Amandi, A., Schiaffino, S. and Campo, M., "Evaluating Bayesian Networks' Precision for Detecting Students' Learning Styles.",Computers & Education, Elsevier., vol. 49,no.3, pp.794-808, 2007

[52] Gomes, A. Santos,A.,Carmo,L. and Mendes, A.J., "Learning styles in an e-learning tool.", International Conference on Engineering Education (ICEE2007), Coimbra, 2007

[53] Gomes, P. and Antunes , B. and Rodrigues, L. and Santos, A. and Barbeira, J. and Carvalho,R., "Using Ontologies for eLearning Personalization.", 3rd E-learning Conference – Computer Science Education, Coimbra, Portugal, 2006

[54] Görgün, İ. Türker, A., Ozan, Y. and Heller,J.,"Learner Modeling To Facilitate Personalized e-Learning experience.", CELDA 2005, pp.231-237, 2005

[55] Graf, S and Kinshuk., "An Approach for Detecting Learning Styles in Learning Management Systems", Proceedings of the International Conference on Advances Learning Technologies (ICALT 06), Kerkrade, Netherlands, pp. 161-163., July 2006

[56] Graf, S and Kinshuk., "Enabling Learning Management Systems to Identify

Learning Styles*.", Proceedings of the International Conference on Interactive Computer Aided Learning (ICL 06), Villach, Austria, September 2006

[57] Graf,S., Viola,S.R. and Kinshuk., "Detecting Learners' Profiles based on the Index of Learning Sstyles Data.", Proceedings of the International Workshop on Intelligent and Adaptive Web-based Educational Systems (IAWES 2007), Hiroshima, Japan, pp. 233-238, 2007

[58] Graf,S., Viola,S.R. and Kinshuk.,"Automatic Student Modelling for Detecting Learning Style Preferences in Learning Management Systems.", Proceedings of the IADIS International Conference on Cognition and Exploratory Learning in Digital Age (CELDA 2007) Algarve, Portugal, pp. 172-179, December 2007.

[59] Gu, Qinyi, and Summer,T., "Support Personalization in Distributed E-Learning Systems through Learner Modeling.",Information and Communication Technologies 2006, ICTTA '06, 2nd, vol.1, pp. 610- 615, 2006

[60] Henze, N., Dolog, P. and Nejdl,W., "Reasoning and Ontologies for Personalized
E-Learning in the Semantic Web. Educational Technology & Society", vol.7, no.4, pp.82-97, 2004

[61] Li, X and Chang,S., "A Personalized E-Learning System Based on User Profile Constructed Using Information Fusion.", The eleventh International Conference on Distributed Multimedia Systems (DMS'05), pp. 109-114 Banff, Canada, 2005

[62] Nabeth, T., Razmerita, L., Angehrn, A. and Roda,C., "InCA: a Cognitive Multi-Agents Architecture for Designing Intelligent & Adaptive Learning Systems.", A special Issue of ComSIS journal, vol.3,no.2,2005

[63] Papanikolaou, K., and Grigoriadou,M.,"Accommodating learning style characteristics in Adaptive Educational Hypermedia.",Workshop on Individual Differences in Adaptive Hypermedia organised in AH2004, Part I, The Eindhoven University of Technology, The Netherlands, pp. 77-86, 2004

[64] Peña, C., Narzo, J. and de la Rosa,J., "Intelligent Agents in a Teaching and Learning Environment on the Web.", Proceedings of the International Conference on Advanced Learning Technologies, 2002

[65] Qiu, F. and Cho,J., "Automatic identification of user interest for personalized

search.", In Proceedings of the 15th International Conference on World Wide Web (Edinburgh, Scotland). WWW '06. ACM Press, New York pp.727-736.,2006

[66] Special issue on statistical and probabilistic methods for user modeling. User Model User-Adap Inter 17, 2007

[67] Speretta, M. and Gauch,S.,"Personalized Search Based on User Search Histories.", Proceedings of the 2005 IEEE/WIC/ACM International Conference on Web Intelligence, pp. 622-628, 2005

[68] Tian, F., Zheng, Q., Gong, Z., Du, J. and Li, R., " Personalized Learning Strategies in an intelligent e-Learning Environment.", The 11th International Conference on Computer Supported Cooperative Work in Design (CSCWD 2007) Melbourne, Australia, pp. 973-978, 2007

[69] Wade, P. and Ashman,H.,"Evolving The Infrastructure For Technology Enhanced Distance Learning.", IEEE Internet Computing, vol.11,no.13, pp.16-18, 2007

[70] Zigoris, P. and Zhang,Y., "Bayesian Adaptive User Profiling with Explicit & Implicit Feedback.", CIKM'06, Arlington, Virginia, USA., 2006

[71] Brusilovsky,P.,"Methods and Techniques of Adaptive Hypermedia. User Model.", User-Adapt. Interact.,vol. 6,no.(2-3), pp. 87-129,1996

[72] "Moodle", retrieved from http://moodle.org/

[73] Felder, R.M.; Soloman, B.A.: Index of Learning Styles Questionnaire. (1997), retrieved from http://www.engr.ncsu.edu/learningstyles/ilsweb.html, last visited on Febuary 2007

[74] Foster, I., Kesselman,C., Nick, J.M. and Tuecke,S., "The Physiology of the Grid An Open Grid Services Architecture for Distributed Systems Integration", Globus Project, 2002 , retrieved from www.globus.org/research/papers/ogsa.pdf , last visited on November 2005

[75] Kohavi,R., "Scaling Up the Accuracy of Naive-Bayes Classifiers: a Decision-Tree Hybrid", Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, pp.202-207, 1996

[76] "OGSA-DAI" , retrieved from http://www.ogsadai.org.uk/about/ogsa-dai, last visited on October 2005

[77] Hong,N.C., "OGSA-DAI The Next Generation", BIODA Workshop Daresbury Laboratory, 26 September 2005

[78] "Distributed Query Processing Using OGSA-DAI", retrieved from http://www.ogsadai.org.uk/about/ogsa-dqp, last visited on October 2005

[79] Alpdemir, M.N., Mukherjee,A. Gounaris,A., Paton,N.W., Watson,P., Fernandes, A.A.A. and Fitzgerald,D.J., "OGSA-DQP: A Service for Distributed Querying on the Grid", EDBT 2004, LNCS 2992, pp. 858–861, Springer-Verlag Berlin Heidelberg, 2004

[80] Watson,P., "Databases in Grid Applications: Locality and Distribution", In Proceedings of the Database:Enterprise,Skills and Innovation 22nd British National Conference on Databases,BNCOD 22,Sunderland,UK,July 5-7 2005

[81] Gounaris,A., "Resource Aware Query Processing On The Grid", PhD Thesis 2005

[82] Van Zwanenberg, N., Wilkinson, L. J., and Anderson, A., "Felder and Silverman's Index of Learning Styles and Honey and Mumford's Learning Styles Questionnaire: How Do They Compare and Do They Predict Academic Performance?", Educational Psychology, vol.20 , no.3, pp.365 - 380, 2000

[83] Viola, S. R., Graf, S., Kinshuk, and Leo, T. , "Investigating Relationships within the Index of Learning Styles: A Data-Driven Approach.",International Journal of Interactive Technology and Smart Education, vol.4,no.1, pp.7-18, 2007

[84] Janssens, D., Wets, G., Brijs, T. , Vanhoof, K., Arentze, T. and Timmermans, H. ,"Integrating Bayesian networks and decision trees in a sequential rule-based transportation model ", European Journal of Operational Research, vol. 175, no.1, pp.16-34, 2006

[85] Friedman,N. and Goldszmidt, M. "Building Classifiers using Bayesian Networks",
Proceedings of the Thirteenth National Conference on Artificial Intelligence and Eighth Innovative Applications of Artificial Intelligence Conference, 13.AAAI 96, 8.IAAI 96 pp. 1277-1284, 1996

[86] Boutell, M.R., Luo,J., Shena,X. and Brown,C.M.,"Learning multi-label scene classification", Elsevier B.V., Pattern Recognition, vol. 37, no. 9, pp. 1757-1771, 2004

[87] Tsoumakas,G. and Katakis,I. "Multi Label Classification: An Overview", International Journal of Data Warehousing and Mining, David Taniar (Ed.), Idea Group Publishing, vol.3,no.3, pp. 1-13, 2007.

[88] Diplaris, S., Tsoumakas, G., Mitkas, P. and Vlahavas, I., "Protein Classification with Multiple Algorithms", Proceedings of the 10th Panhellenic Conference on Informatics (PCI 2005), Volos, Greece, November 2005

[89] Goncalves, T. and Quaresma, P. , "A Preliminary Approach to the Multilabel Classification Problem of Portuguese Juridical Documents", Proceedings of the 11th Portuguese Conference on Artificial Intelligence, 2003.

[90] Lauser, B. and Hotho, A. , "Automatic multi-label subject indexing in a multilingual environment", Proceedings of the 7th European Conference in Research and Advanced Technology for Digital Libraries, 2003.

[91] Li, T. and Ogihara, M., "Detecting emotion in music", Proceedings of the International Symposium on Music Information Retrieval, Washington D.C., USA, 2003

[92] Clare, A. and King, R.D. , "Knowledge Discovery in Multi-Label Phenotype Data", Proceedings of the 5th European Conference on Principles of Data Mining and Knowledge Discovery (PKDD 2001), Freiburg, Germany, 2001

[93] Freund, Y. and Schapire, R.E., "A decision-theoretic generalization of on-line learning and an application to boosting", Journal of Computer and System Sciences, vol. 55, no. 1, pp. 119-39, 1997

[94] Schapire, R.E. and Singer, Y. ,"Boostexter: a boosting-based system for text categorization", Machine Learning, vol. 39, no. (2-3), pp. 135-68, 2000

[95] Zhang, M.-L. and Zhou, Z.-H., "A k-Nearest Neighbor Based Algorithm for Multi-label Classification", Proceedings of the 1st IEEE International Conference on Granular Computing, 2005

[96] McCallum, A., "Multi-label text classification with a mixture model trained by EM", Proceedings of the AAAI' 99 Workshop on Text Learning, 1999

[97] Elisseeff, A. and Weston, J., "A kernel method for multi-labelled classification", Advances in Neural Information Processing Systems vol.14, 2002

[98] Godbole, S. and Sarawagi, S., "Discriminative Methods for Multi-labeled Classification", Proceedings of the 8th Pacific-Asia Conference on Knowledge Discovery and Data Mining, 2004.

[99] Thabtah, F.A., Cowling,P. and Peng,Y., "MMAC:A new Multi-class, Multi-label Associative Classification Approach", Proceedings of the 4th IEEE International Conference on Data Mining, 2004

[100] "SCO", retrieved from http://www.eduworks.com/LOTT/Tutorial/scormconcepts.html, last visited on May 2006

[101]"PoSTech!", retrieved from http://www.postech.me.uk/

[102] Seyedarabi,F., Peterson,D. & Keenoy,K., "Personalised Search Tool for teachers – PoSTech!", I-manager's Journal of Educational Technology, Teacher's Use of Technology For Creative Learning Environment, vol.1,no.2, pp.38-49,2005

[103] Seyedarabi,F.,"The Missing Link: How search engines can support the informational needs of teachers.", ACM Magazine, 6 April 2006

[104] Seyedarabi, F., "Teaching Teachers to Google: Development of a Personalised Search Tool for Teachers' online searching (research in progress).", In K. McFerrin et al. (Eds.), Proceedings of Society for Information Technology and Teacher Education International Conference, pp. 3414-3419, Chesapeake, VA: AACE, 2008.

[105] Graf, S., Viola, S. R., Kinshuk and Leo, T., " In-depth analysis of the Felder-Silverman learning style dimensions.",Journal of Research on Technology in Education, vol.40,no.1, pp.79–93, 2007.

[106] Tsoumakas, G., and Vlahavas, I., "Random k-labelsets: An ensemble method for multilabel classification.", In Proceedings of the 18th European conference on machine learning (ECML 2007) pp. 406–417,2007.

[107] Petcu,D., Macariu,G., Francu,M. and Carstea,A., "Grid services vs. Web services", retrieved from
http://www.symbolic-computation.org/science/images/e/ec/Dana-Petcu_GridServicesAndWebServices.pdf, last visited on May 2009

APPENDIX A: HOW TO INSTALL OUR SYSTEM?

Prerequisite software

1. Install j2sdk 1.4.2_07 and set environment variable JAVA_HOME to its directory and add its "bin" directory to the system environment variable named "Path".
2. Install JAKARTA ANT 1.6 and set environment variable ANT_HOME to its directory.Then its "bin" directory to the system environment variable named "Path".
3. Install Globus Toolkit 4.0.1 (GT4) Java Web Services Core and set environment variable GLOBUS_LOCATION to its directory.
4. Install jakarta Tomcat5.0.28 web server and set environment variable CATALINA_HOME to its directory.
5. Install MySQL 5.0.15 database management system

Prerequisite JAR's

1. The following jar files are needed only if you are interested to use the associated functionality.

OGSA-DAI Functionality	JAR	Package Name	URL
File-related activities	jakarta-oro-2.0.8.jar	Jakarta Oro 2.0.8	http://jakarta.apache.org/oro
XML database-related activities	One of: • xmldb.jar • xmldb-api.jar • xmldb-api-20030701.jar	Xindice 1.0	Available in Xindice - http://xml.apache.org/xindice
Indexed file-related activities	lucene-1.4.3.jar	Lucene 1.4.3	http://jakarta.apache.org/lucene

2. For MySQL database the database driver namely mysql-connector-java-3.1.8-bin.jar is required.

PART1: Deploy OGSADAI-WSRF2.1 binary distribution

1. Install OGSADAI-WSRF2.1 and put prerequested JAR files into the "lib" directory.
2. Deploy GT4 onto Tomcat by the following commands from the command line:

```
$ cd path/to/Globus/directory
$ ant -f share/globus_wsrf_common/tomcat/tomcat.xml deployTomcat
  -Dtomcat.dir=$CATALINA_HOME
$ cd path/to/OGSA-DAI/binary/directory
```

C:\EBRU\ws-core-4.0.1>ant -f share/globus_wsrf_common/tomcat/tomcat.xml deployTo
mcat -Dtomcat.dir=%CATALINA_HOME%
Buildfile: share\globus_wsrf_common\tomcat\tomcat.xml

deployTomcat:

_baseTomcatDeploy:
[mkdir] Created dir: C:\EBRU\TOMCAT\webapps\wsrf
[copy] Copying 90 files to C:\EBRU\TOMCAT\webapps\wsrf
[mkdir] Created dir: C:\EBRU\TOMCAT\webapps\wsrf\WEB-INF
[copy] Copying 30 files to C:\EBRU\TOMCAT\webapps\wsrf\WEB-INF
[mkdir] Created dir: C:\EBRU\TOMCAT\webapps\wsrf\WEB-INF\classes
[copy] Copying 1 file to C:\EBRU\TOMCAT\webapps\wsrf\WEB-INF\classes
[copy] Copying 1 file to C:\EBRU\TOMCAT\webapps\wsrf\WEB-INF\classes
[mkdir] Created dir: C:\EBRU\TOMCAT\webapps\wsrf\WEB-INF\lib
[copy] Copying 1 file to C:\EBRU\TOMCAT\webapps\wsrf\WEB-INF
[copy] Copying 1 file to C:\EBRU\TOMCAT\common\endorsed
[copy] Copying 58 files to C:\EBRU\TOMCAT\webapps\wsrf\WEB-INF\lib

BUILD SUCCESSFUL
Total time: 24 seconds
C:\EBRU\ws-core-4.0.1>cd c:\ebru\ogsadai-wsrf-2.1

C:\EBRU\ogsadai-wsrf-2.1>

3. Deploy OGSA-DAI WSRF onto Tomcat using a GUI by the following commands from the command line:

```
$ cd path/to/OGSA-DAI/binary/directory
$ ant guiDeployTomcat
```

- Select your Tomcat distribution by pressing the ***Select...*** button.
- If ***Path*** is okey,then press *Next.*
- You now can select any of the prerequisite JAR files (as listed in the table above) **only if** you intend to use the functionality that the JARs support. When you are happy with your selections, press *Next.*

4. You will now be requested to press *Next* so that the tool can deploy OGSA-DAI WSRF. Note that you may get some warnings of the type:

```
[copy] Warning: Could not find file
/some/path/${dai.build.jars.0} to copy.
```

These can safely be ignored.

5. When the deployment has completed a completion window will be displayed. Press ***Finish*** to exit the tool.

C:\EBRU\ogsadai-wsrf-2.1>ant guiDeployTomcat

Buildfile: build.xml

guiTomcatHelper:

guiDeployTomcat:

wizardHelper:

BUILD SUCCESSFUL

Total time: 12 minutes 4 seconds

C:\EBRU\ogsadai-wsrf-2.1>

PART2: Exposing Data Resources using OGSA-DAI WSRF

1. Deploying data service using GUI

```
$ ant guiDeployDataServiceTomcat
```

- Select your Tomcat distribution by pressing the ***Select...*** button.

- If ***Path*** is okey,then press *Next.*
- Enter the the **Service Path** field. For example "`ogsadai/DataService`" specifies a data service whose URL will be http://HOST:PORT/wsrf/services/ogsadai/DataService.
- Clicking on the **Dynamically configurable?** box requests that the data service support dynamic service configuration.(Optional)
- Specify the support for concurrent requests that each resource exposed by the service will provide.

Maximum Requests specifies the number of requests that each resource can run concurrently. The default is 10.

Request Queue Length specifies the number of requests that each resource can queue if its maximum number of concurrent requests has been reached. Once an existing request completes the next request in the queue will start executing. The default is 5.

- Press *Next* when you entered this information.
- Press *Next* so that the tool can deploy the data service. If any problems occur then the tool will notify you of these.
- When the deployment has completed a completion window will be displayed. Press ***Finish*** to exit the tool.
- You will need to shutdown and restart Tomcat before clients are able to access the new data service.

2. Checking the OGSA-DAI Data Service Deployment

- Restart Tomcat.
- Using a Web browser, visit the following URL: http://TOMCAT.HOST:TOMCAT.PORT/wsrf/services
- A list of services will be displayed. The one corresponding to deployed should be seen,e.g.

```
ogsadai/DataService (wsdl)
```

- When clicking on the (wsdl) link and the WSDL describing the service should appear;e.g.

```
<?xml version="1.0" encoding="UTF-8" ?>
<wsdl:definitions name="DataService"
...
```

- While Tomcat is running, test the data service resource using the ListResources client by running the following command from within the OGSA-DAI WSRF binary distribution directory:

```
$ ant listResourcesClient -Ddai.url=SERVICE-URL
```

where `-Ddai.url=` specifies the URL of the data service. If omitted then a default URL of `http://localhost:8080/wsrf/services/ogsadai/DataService` is used.

`http://localhost:8090/wsrf/services/ogsadai/service1`

The data service resources exposed by your service will be displayed; e.g.

```
[java] Contacting ...
http://localhost:8080/wsrf/services/ogsadai/DataService
[java] Service version: OGSA-DAI WSRF 2.1
[java] Number of resources: 2
[java] Resource: MySQLResource
[java] Resource: AnotherResource
```

C:\EBRU\ogsadai-wsrf-2.1>ant listResourcesClient -Ddai.url=http://localhost:8080

/wsrf/services/ogsadai/DataService

Buildfile: build.xml

setupClientSecurity:

listResourcesClient:

[java] Service version: OGSA-DAI WSRF 2.1

[java] Number of resources: 0

BUILD SUCCESSFUL

Total time: 18 seconds

C:\EBRU\ogsadai-wsrf-2.1>

3. Create a new data service resource file and deploy data service resource

- Restart Tomcat
- Run the following command from within the OGSA-DAI WSRF binary distribution directory:

$ ant guiCreateResourceTomcat

- Select your database product from the offered ones and press ***Next***
- Now, you have to enter information about the data service resource and your database product. This information includes:

Data Resource URI - data resource URI (required). This must be compatible with the driver class specified next. E.g. jdbc:mysql://myhost.example.com:3306/DATABASE_NAME

Data Resource Driver Class - data resource driver class name. This is optional only if you selected a **Files and directories** product on the previous window. E.g. org.gjt.mm.mysql.Driver

Vendor - data resource product vendor (optional). E.g. MySQL

Version - data resource product version (optional). E.g. 5.0.15

Database access credential - Grid certificate credentials of a user permitted to access the data resource. If omitted then any user will be allowed access. E.g. left empty

Database user ID - data resource user name. Optional only if there is no user name required for a database. E.g. ogsadai

Database password - corresponding data resource password. Optional if there is no user name required, or if the password is null. E.g. ogsadai

Data Service Resource ID - name for the data service resource (required). E.g. MySQLResource

- Then, press **Next**.

- Select the JARs that implement your database driver if these have not already been installed on the server.
- Enter a file name in which to save your values. These will be saved in a new data service resource file. Press the ***Select...*** button, enter a new file name in the **File name** field then press ***Select.***
- If ***Path*** is okey,then press ***Next.***
- Press ***Next*** if you want to proceed to deploy your data service resource.
- Select your Tomcat distribution by pressing the ***Select...*** button.
- If ***Path*** is okey,then press ***Next*** **.**
- Press ***Next*** so that the tool can deploy the data service resource. If any problems occur then the tool will notify you of these.
- When the deployment has completed a completion window will be displayed. Press ***Finish*** to exit the tool.

4. Add data service resource to data service

- Under Tomcat run the following command from within the OGSA-DAI WSRF binary distribution directory:

  $ ant guiAddResourceTomcat

- Select your Tomcat distribution by pressing the ***Select...*** button.
- If ***Path*** is okey,then press ***Next*** **.**
- Enter

  - The local URL of the data service, for example `ogsadai/DataService`, in the **Service Path** field.

- The ID of the data service resource, for example `MyRelationalResource` in the **Data Service Resource ID** field.

- Then, press Next.
- Press Next so that the tool can add the data service resource to the data service. If any problems occur then the tool will notify you of these.
- When the addition has completed a completion window will be displayed. Press Finish to exit the tool.
- You will need to shutdown and restart Tomcat before clients are able to access the new data service resource via the data service.

5. Restart Tomcat

6. Test whether the new data service resource was successfully added using the OGSA-DAI WSRF `listResourcesClient` client by running the following command from within the OGSA-DAI WSRF binary distribution directory:

```
$ ant listResourcesClient -Ddai.url=SERVICE-URL
```

where `-Ddai.url=` specifies the URL of the data service. If omitted then a default URL of `http://localhost:8080/wsrf/services/ogsadai/DataService` is used.

C:\EBRU\ogsadai-wsrf-2.1>ant listResourcesClient

Buildfile: build.xml

setupClientSecurity:

listResourcesClient:

[java] Service version: OGSA-DAI WSRF 2.1

[java] Number of resources: 1

[java] Resource: MySQLResource

BUILD SUCCESSFUL

Total time: 18 seconds

C:\EBRU\ogsadai-wsrf-2.1>

PART 3. Deploy OGSA-DQP 3.0 Source Distribution

1. Installing OGSA-DQP metadata extractors on OGSA-DAI data service resources (required to wrap databases that we want to query)

- Edit the `dataResourceConfig.xml` file of the MySQL data service resource.

"dataResourceConfig.xml" file can be found under Tomcat in the directory "webapps/wsrf/WEB-INF/etc/ogsadai_wsrf/RESOURCE-ID" where RESOURCE-ID is the resource ID of the OGSA-DAI data service resource in question.

In this file, change the callback attribute of the <databaseSchema> element (which is a child of <relationalMetaData>) to:

```
uk.org.ogsadai.dqp.dataresource.DQPMySQLMetadataExtractor
```

- Deploy the MySQL metadata extractor class

Execute the following in the metadata_extractor directory of the OGSA-DQP distribution:

```
$ ant deploy.tomcat.wsrf
```

This will make the metadata extractor available to OGSA-DAI.

C:\EBRU\ogsadqp\metadata_extractor>ant deploy.tomcat.wsrf

Buildfile: build.xml

prepare:

[echo] setting up build directories...

[mkdir] Created dir: C:\EBRU\ogsadqp\metadata_extractor\build

[mkdir] Created dir: C:\EBRU\ogsadqp\metadata_extractor\build\classes

[mkdir] Created dir: C:\EBRU\ogsadqp\metadata_extractor\ext

[mkdir] Created dir: C:\EBRU\ogsadqp\metadata_extractor\lib

[copy] Copying 3 files to C:\EBRU\ogsadqp\metadata_extractor\ext

compile:

[javac] Compiling 1 source file to C:\EBRU\ogsadqp\metadata_extractor\build\ classes

[echo] compilation of metadata extractor completed.

jar:

[echo] building the md extractor jar file

[jar] Building jar: C:\EBRU\ogsadqp\metadata_extractor\lib\mdextractor.jar

[echo] built metadata extractor jar file

deploy.tomcat.wsrf:

[copy] Copying 1 file to C:\EBRU\TOMCAT\webapps\wsrf\WEB-INF\lib

BUILD SUCCESSFUL

Total time: 11 seconds

C:\EBRU\ogsadqp\metadata_extractor>

2. Installing the OGSA-DQP Query Evaluation Service (Evaluator installation)

- Run the following from within the `evaluator` directory of the OGSA-DQP distribution:

```
$ ant deploy
```

This will create a WAR file for the QES and deploy it on Tomcat.

C:\EBRU\ogsadqp\evaluator>ant deploy

Buildfile: build.xml

axisver:

cleanext:

prepare:

[echo] setting up build directories...

[mkdir] Created dir: C:\EBRU\ogsadqp\evaluator\build

[mkdir] Created dir: C:\EBRU\ogsadqp\evaluator\ext

[mkdir] Created dir: C:\EBRU\ogsadqp\evaluator\build\webapp

[mkdir] Created dir: C:\EBRU\ogsadqp\evaluator\build\generated

[mkdir] Created dir: C:\EBRU\ogsadqp\evaluator\build\generated\stubs

[mkdir] Created dir: C:\EBRU\ogsadqp\evaluator\build\classes

[mkdir] Created dir: C:\EBRU\ogsadqp\evaluator\build\lib

[mkdir] Created dir: C:\EBRU\ogsadqp\evaluator\build\schema

[echo] Copying all the schema files...

[copy] Copying 4 files to C:\EBRU\ogsadqp\evaluator\build\schema

[copy] Copying 10 files to C:\EBRU\ogsadqp\evaluator\build\schema

[echo] Copying all the common libraries...

[copy] Copying 7 files to C:\EBRU\ogsadqp\evaluator\ext

copy.axis-1.2-rc2:

[echo] Copying Axis-1.2-RC2 libraries

[copy] Copying 14 files to C:\EBRU\ogsadqp\evaluator\ext

[copy] Copying 1 file to C:\EBRU\ogsadqp\evaluator\etc

copy.axis-1.2.1:

stubs:

[echo] Generating service stubs/skeletons for Evaluator

[echo] Done generating stubs for the evaluator...

```
compile.stubs:
  [javac] Compiling 69 source files to C:\EBRU\ogsadqp\evaluator\build\classes
   [echo] compilation of evaluator completed.
genPart:
   [echo] generating partition code for evaluator...
   [java] -- Suppressing non fatal warnings.
   [echo] done generating partition code.
genProgMon:
   [echo] generating progress monitoring information code...
   [java] -- Suppressing non fatal warnings.
   [echo] done generating progress monitoring information code.
compile:
  [javac] Compiling 171 source files to C:\EBRU\ogsadqp\evaluator\build\classes
   [echo] compilation of evaluator completed.
jar:
   [echo] building jar file for Evaluator...
    [jar] Building jar: C:\EBRU\ogsadqp\evaluator\build\lib\dqp-evaluator.jar
   [echo] jar file for Evaluator built.
war:
    [war] Building war: C:\EBRU\ogsadqp\evaluator\build\webapp\dqp-
evaluator.war
deploy:
   [copy] Copying 1 file to C:\EBRU\TOMCAT\webapps
   [echo] Please restart tomcat for full deployment of the QueryEvaluationService
```

BUILD SUCCESSFUL

Total time: 1 minute 36 seconds

C:\EBRU\ogsadqp\evaluator>

- Startup Tomcat. If the above step is performed successfully, the result is that the QES is deployed on Tomcat, and you should be able to view the WSDL at:

http://<host>:<port>/dqp-evaluator/services/QueryEvaluationService?wsdl

e.g., http://localhost:8080/dqp-evaluator/services/QueryEvaluationService?wsdl

3. Installing the OGSA-DQP 3.0 GDQS (Coordinator installation)

ATTENTION! : Before starting the installation of GDQS, the new version of "DQPResultsToWebRowSet.java" should overwrite the old version of this file, which can be found in the directory:

```
OGSADQP/coordinator/src/uk/org/ogsadai/dqp/common/utils
```

where OGSADQP is the directory in which you have unpacked the OGSA-DQP download.

- Install Polar* query compiler.

The directory polarstar/cygwin-binary contains a compiled Polar* executable (qcomp.exe) for the convenience of Windows users.

In order to use this, it is necessary to obtain the Cygwin 1.5.18 DLL by downloading Cygwin and following the Cygwin installation instructions. (Cygwin 1.5.19 DLL, the only currently available Cygwin is already downloaded and cygwin1.dll is given in setup files☺)

Copy the file "cygwin1.dll" from Cygwin to the directory polarstar/cygwin-binary. When deploying the coordinator, specify the absolute filename of qcomp.exe as the location of Polar*.

- Deploy an OGSA-DAI data service as described in Part2 step1. The data service initially exposes 0 data service resources.(e.g. data service name can be "ogsadai/service1").

e.g;

C:\PhD\ogsadai-wsrf-2.1>ant listResourcesClient -Ddai.url=http://localhost:8090/

wsrf/services/ogsadai/service1

Buildfile: build.xml

setupClientSecurity:

listResourcesClient:

[java] Service version: OGSA-DAI WSRF 2.1

[java] Number of resources: 0

BUILD SUCCESSFUL

Total time: 4 seconds

C:\PhD\ogsadai-wsrf-2.1>

C:\PhD\ogsadai-wsrf-2.1>

- Run the following from within the coordinator directory of the OGSA-DQP distribution:

```
$ ant deploy.tomcat.wsrf -Dresource.id=RESOURCE-ID
 -Dservice.path=SERVICE-PATH -Dservice.url=SERVICE-URL
 -Dpolarstar=POLARSTAR-EXECUTABLE
```

Where:

- `RESOURCE-ID` is the resource ID of the GDQS factory data service resource. Typically this argument will take the value "dqp-factory".
- `SERVICE-PATH` is the path of the data service deployed in step (3.2), for example: ogsadai/service1.

- `SERVICE-URL` is the URL of the deployed data service, for example: `"http://rimside.ncl.ac.uk:8080/wsrf/services/ogsadai/service1"`
- `POLARSTAR-EXECUTABLE` is the name of the prerequisite Polar* executable (typically "qcomp" or "qcomp.exe").

[**Warning**: please be accurate when specifying the `resource.id`, `service.url` and `polarstar` arguments. The consequences of a mistake made here will cause the coordinator to fail with a rather generic error message during query compilation or execution, however no error message will be given at the time that the above command is executed.]

C:\PhD>cd ogsadqp

C:\PhD\ogsadqp>cd coordinator

C:\PhD\ogsadqp\coordinator>ant deploy.tomcat.wsrf -Dresource.id=dqp-factory -Dse

rvice.path=ogsadai/service1 -Dservice.url=http://localhost:8090/wsrf/services/og

sadai/service1 -Dpolarstar=C:\PhD\ogsadqp\polarstar\cygwin-binary\qcomp.exe

Buildfile: build.xml

cleanext:

prepare:

[echo] setting up build directories...

[mkdir] Created dir: C:\PhD\ogsadqp\coordinator\build

[mkdir] Created dir: C:\PhD\ogsadqp\coordinator\build\classes

[mkdir] Created dir: C:\PhD\ogsadqp\coordinator\ext

[mkdir] Created dir: C:\PhD\ogsadqp\coordinator\build\generated

[mkdir] Created dir: C:\PhD\ogsadqp\coordinator\build\lib

[copy] Copying 13 files to C:\PhD\ogsadqp\coordinator\ext

```
generateconfiguration:
    [echo] generating gdqs configuration code...
    [java] -- Suppressing non fatal warnings.
    [echo] done generating computational resource metadata code.
compile.generated:
   [javac] Compiling 28 source files to C:\PhD\ogsadqp\coordinator\build\classes
    [echo] compilation of castor generated code completed.
jar.evaluator.wsrf:
axisver:
cleanext:
  [delete] Deleting directory C:\PhD\ogsadqp\evaluator\ext
prepare:
    [echo] setting up build directories...
   [mkdir] Created dir: C:\PhD\ogsadqp\evaluator\ext
    [echo] Copying all the schema files...
    [copy] Copying 4 files to C:\PhD\ogsadqp\evaluator\build\schema
    [copy] Copying 10 files to C:\PhD\ogsadqp\evaluator\build\schema
    [echo] Copying all the common libraries...
    [copy] Copying 7 files to C:\PhD\ogsadqp\evaluator\ext
copy.axis-1.2-rc2:
    [echo] Copying Axis-1.2-RC2 libraries
    [copy] Copying 14 files to C:\PhD\ogsadqp\evaluator\ext
    [copy] Copying 1 file to C:\PhD\ogsadqp\evaluator\etc
copy.axis-1.2.1:
```

stubs:

[echo] Generating service stubs/skeletons for Evaluator

[echo] Done generating stubs for the evaluator...

compile.stubs:

[javac] Compiling 69 source files to C:\PhD\ogsadqp\evaluator\build\classes

[echo] compilation of evaluator completed.

genPart:

[echo] generating partition code for evaluator...

[java] -- Suppressing non fatal warnings.

[echo] done generating partition code.

genProgMon:

[echo] generating progress monitoring information code...

[java] -- Suppressing non fatal warnings.

[echo] done generating progress monitoring information code.

compile:

[javac] Compiling 171 source files to C:\PhD\ogsadqp\evaluator\build\classes

[echo] compilation of evaluator completed.

jar:

[echo] building jar file for Evaluator...

[jar] Building jar: C:\PhD\ogsadqp\evaluator\build\lib\dqp-evaluator.jar

[echo] jar file for Evaluator built.

[copy] Copying 1 file to C:\PhD\ogsadqp\coordinator\ext

compile.coordinator.wsrf:

[javac] Compiling 34 source files to C:\PhD\ogsadqp\coordinator\build\classes

[javac] Note: C:\PhD\ogsadqp\coordinator\src\uk\org\ogsadai\dqp\gdqs\OQLQueryStatementActivity.java uses or overrides a deprecated API.

[javac] Note: Recompile with -deprecation for details.

[echo] compilation of coordinator completed.

jar.gdqs.wsrf:

[echo] building gdqs jar file

[copy] Copying 2 files to C:\PhD\ogsadqp\coordinator\build\classes\uk\org\o

gsadai\dqp\gdqs

[jar] Building jar: C:\PhD\ogsadqp\coordinator\build\lib\gdqs.jar

[echo] built GDQS jar file

deploy.tomcat.wsrf:

[echo] Data service resource added!

[copy] Copying 1 file to C:\PhD\TOMCAT\webapps\wsrf\share\schema\ogsadai\xs

d\activities

[copy] Copying 1 file to C:\PhD\TOMCAT\webapps\wsrf\share\schema\ogsadai\xs

d\activities

[copy] Copying 4 files to C:\PhD\TOMCAT\webapps\wsrf\WEB-INF\lib

[mkdir] Created dir: C:\PhD\TOMCAT\webapps\wsrf\WEB-INF\etc\ogsadai_wsrf\dqp

-factory

[copy] Copying 3 files to C:\PhD\TOMCAT\webapps\wsrf\WEB-INF\etc\ogsadai_ws

rf\dqp-factory

[mkdir] Created dir: C:\PhD\TOMCAT\webapps\wsrf\WEB-INF\etc\ogsadai_wsrf\dqp

-factory\instance

[copy] Copying 5 files to C:\PhD\TOMCAT\webapps\wsrf\WEB-INF\etc\ogsadai_ws

rf\dqp-factory\instance

[echo] deployment of coordinator completed.

[copy] Copying 1 file to C:\PhD\TOMCAT\webapps\wsrf\WEB-INF\classes

BUILD SUCCESSFUL

Total time: 23 seconds

C:\PhD\ogsadqp\coordinator>

Start-up Tomcat. If these steps are performed successfully, the result is that the GDQS factory data service resource is added to the data service resources exposed by the specified data service.

C:\PhD\ogsadai-wsrf-2.1>ant listResourcesClient -Ddai.url=http://localhost:8090/

wsrf/services/ogsadai/service1

Buildfile: build.xml

setupClientSecurity:

listResourcesClient:

[java] Service version: OGSA-DAI WSRF 2.1

[java] Number of resources: 1

[java] Resource: dqp-factory

BUILD SUCCESSFUL

Total time: 3 seconds

C:\PhD\ogsadai-wsrf-2.1>

4. Configuration (done only on coordinator side!!)

- Configuring OGSA-DQP involves the following:

- Specifying the evaluators which can be used by a GDQS data service resource.
- Specifying the OGSA-DAI data service resources (which wrap relational databases) that can be used when a client interacts with a GDQS data service resource.
- Specifying the analysis (Web) services that can be used when a client interacts with a GDQS data service resource.

The configuration file is the `dqp-factory-configuration.xml` file which can be found under Tomcat in `webapps/wsrf/WEB-INF/etc/ogsadai_wsrf/instance.`

Note that you have to shutdown Tomcat (if it is running) before editing this file, and restart Tomcat once you have saved the file. The factory configuration is basically a set of default configuration parameters what will be used to create any GDQS data service resources.

Here is an example dqp-factory-configuration.xml file:

```
<?xml version="1.0" encoding="utf-8" ?>
<DQPConfiguration xmlns="http://uk.org.ogsadai/dqp/configuration">
   <DQPEvaluatorList>
      <EvaluatorURI>http://144.122.166.94:8080/dqp-
evaluator/services/QueryEvaluationService</EvaluatorURI>
      <EvaluatorURI>http://144.122.167.11:8080/dqp-
evaluator/services/QueryEvaluationService</EvaluatorURI>
   </DQPEvaluatorList>
      <DataResourceList>
       <ImportedDataSource>

<URI>http://144.122.166.94:8080/wsrf/services/ogsadai/DataService1</
URI>
           <ResourceID>MySQLResource</ResourceID>
       </ImportedDataSource>
       <ImportedDataSource>

<URI>http://144.122.167.11:8080/wsrf/services/ogsadai/DataService</U
RI>
           <ResourceID>MySQLResource</ResourceID>
       </ImportedDataSource>
      </DataResourceList>
</DQPConfiguration>
```

- Configuration Constraints

The following constraints are applied when a GDQS factory data service resource attempts to create a GDQS data service resource:

1. At least one evaluator service must be specified.
2. At least one OGSA-DAI data service resource wrapping a relational database must be specified.
3. It is not possible to use multiple evaluators which are deployed on the same host (IP address). Only one evaluator per host may be specified.
4. There can be no more than one OGSA-DAI data service resource imported from any one host.
5. For each OGSA-DAI data service resource to be imported, an evaluator service deployed on the same host must be specified.

Failure to adhere to these constraints will result in an error when attempting to create a GDQS data service resource using the GDQS factory.

Note that installing the GDQS on the same host as one of the evaluators is a valid configuration.

PART 4. Getting Logs Locally in a File

In order to get logs in a file locally, edit the log4j.properties file under "C:\EBRU2\TOMCAT\webapps\wsrf\WEB-INF\classes" directory on ach machine in the DQP system. Here is a sample log4j.properties file where bold italic parts show the adds for getting logs locally to a file:

```
# Set root category priority to WARN and its only appender to A1.

log4j.rootCategory=ERROR, A1, LOGFILE

# A1 is set to be a ConsoleAppender.

log4j.appender.A1=org.apache.log4j.ConsoleAppender

# A1 uses PatternLayout.
log4j.appender.A1.layout=org.apache.log4j.PatternLayout
log4j.appender.A1.layout.ConversionPattern=%d{ISO8601} %-5p %c{2}
[%t,%M:%L] %m%n
```

```
# LOGFILE is set to be a File appender using a PatternLayout.
log4j.appender.LOGFILE=org.apache.log4j.FileAppender
log4j.appender.LOGFILE.File=ogsa_log.txt
log4j.appender.LOGFILE.Append=true
#log4j.appender.LOGFILE.Threshold=INFO
log4j.appender.LOGFILE.layout=org.apache.log4j.PatternLayout
#log4j.appender.LOGFILE.layout.ConversionPattern=%-4r [%t] %-5p %c
%x - %m%n
#log4j.appender.LOGFILE.layout.ConversionPattern=%d %-5p [%t]
%17c{2}(%13F:%L)  %3x - %m%n
log4j.appender.LOGFILE.layout.ConversionPattern=%d{ISO8601} %-5p
%c{2} [%t,%M:%L] %m%n


# Display any warnings generated by our code
log4j.category.org.globus=INFO

# Comment out the line below if you want to log every authorization
# decision the container makes.
log4j.category.org.globus.wsrf.impl.security.authorization.ServiceAu
thorizationChain=WARN

# Enable SOAP Message Logging
#
log4j.category.org.globus.wsrf.handlers.MessageLoggingHandler=DEBUG

# Uncomment the following line to enable MDS debugging
# log4j.category.org.globus.mds=DEBUG

# Uncomment the following line to enable GRAM debugging
# log4j.category.org.globus.exec=DEBUG

# Uncomment the following line to enable RFT debugging
# log4j.category.org.globus.transfer=DEBUG
```

PART 5. Checklist when Encountering Problems with OGSA-DQP 3.0

Actually log files are the best hint resources to discover the reason of any encountered problem. The followings are the solutions for the common encountered problems concluded during studies.

- When starting GUI on command line, if it fails to open or by switching the user mode, then check the current java version from command line by writing the following command: *java -version*
- When you live such a condition that you state that OGSA-DQP GUI was working last time but not at the moment then the reason can be the dynamic IPs of the machines in the laboratory. You have to check the machine IPs from the

command line by the command:*ipconfig.* Then update the IPs in the dqp-factory-configuration.xml file of GDQS instance(s) if there is a change in IPs.

1) Be sure that the config.xml files of GDQS instances have in URL's "the actual IP" of the machine instead theterm "localhost"
2) Be sure that the config.xml files of GDQS instances have the correct full path of "qcomp.exe" under qcomp tag;e.g.;
<qcomp>C:\EBRU2\ogsadqp\polarstar\cygwin-binary\qcomp.exe</qcomp>

- When submitting your query from GUI and encountering the following message on GUI :

```
uk.org.ogsadai.client.toolkit.exception.DataFormatException:
Data is not in an expected format: WebRowSetToResultSet
```

There can be two reasons:

1) Be sure that you have done what is prompted in Attention part of GDQS installtion given in 3.item of PART3.

If this is the case then do the Attention part and execute the following from the command line in OGSADQP/coordinator directory:

$ ant jar.gdqs.wsrf

This will rebuild the file OGSADQP/coordinator/build/lib/gdqs.jar.

Now shut down Tomcat and copy this file to the following location (you should overwrite the previous version of the file when doing this):

TOMCATwebapps/wsrf/WEB-INF/lib/

2) Your query expression is in a wrong format.Retry again. For the correct syntax,you can get help from the dqp-grammar.txt under OGSADQP/doc directory.

APPENDIX B

Index of Learning Styles Questionnaire

Barbara A. Soloman
First-Year College
North Carolina State University
Raleigh, North Carolina 27695

Richard M. Felder
Department of Chemical Engineering
North Carolina State University
Raleigh, NC 27695-7905

Directions

Please provide us with your full name. Your name will be printed on the information that is returned to you.

Full Name
Ebru Özpolat

For each of the 44 questions below select either "a" or "b" to indicate your answer. Please choose only one answer for each question. If both "a" and "b" seem to apply to you, choose the one that applies more frequently. When you are finished selecting answers to each question please select the submit button at the end of the form.

1. I understand something better after I
   - ○ (a) try it out.
   - ◉ (b) think it through.
2. I would rather be considered
   - ◉ (a) realistic.
   - ○ (b) innovative.
3. When I think about what I did yesterday, I am most likely to get
   - ◉ (a) a picture.
   - ○ (b) words.
4. I tend to
   - ○ (a) understand details of a subject but may be fuzzy about its overall structure.
   - ◉ (b) understand the overall structure but may be fuzzy about details.
5. When I am learning something new, it helps me to

○ (a) talk about it.
◉ (b) think about it.

6. If I were a teacher, I would rather teach a course
   ◉ (a) that deals with facts and real life situations.
   ○ (b) that deals with ideas and theories.

7. I prefer to get new information in
   ◉ (a) pictures, diagrams, graphs, or maps.
   ○ (b) written directions or verbal information.

8. Once I understand
   ○ (a) all the parts, I understand the whole thing.
   ◉ (b) the whole thing, I see how the parts fit.

9. In a study group working on difficult material, I am more likely to
   ○ (a) jump in and contribute ideas.
   ◉ (b) sit back and listen.

10. I find it easier
    ◉ (a) to learn facts.
    ○ (b) to learn concepts.

11. In a book with lots of pictures and charts, I am likely to
    ◉ (a) look over the pictures and charts carefully.
    ○ (b) focus on the written text.

12. When I solve math problems
    ○ (a) I usually work my way to the solutions one step at a time.
    ◉ (b) I often just see the solutions but then have to struggle to figure out the steps to get to them.

13. In classes I have taken
    ◉ (a) I have usually gotten to know many of the students.
    ○ (b) I have rarely gotten to know many of the students.

14. In reading nonfiction, I prefer
    ○ (a) something that teaches me new facts or tells me how to do something.
    ◉ (b) something that gives me new ideas to think about.

15. I like teachers
    ◉ (a) who put a lot of diagrams on the board.
    ○ (b) who spend a lot of time explaining.

16. When I'm analyzing a story or a novel
    ◉ (a) I think of the incidents and try to put them together to figure out the themes.
    ○ (b) I just know what the themes are when I finish reading and then I have to go back and find the incidents that demonstrate them.

17. When I start a homework problem, I am more likely to
    ◉ (a) start working on the solution immediately.

○ **(b)** try to fully understand the problem first.

18. I prefer the idea of
   ◉ **(a)** certainty.
   ○ **(b)** theory.

19. I remember best
   ◉ **(a)** what I see.
   ○ **(b)** what I hear.

20. It is more important to me that an instructor
   ○ **(a)** lay out the material in clear sequential steps.
   ◉ **(b)** give me an overall picture and relate the material to other subjects.

21. I prefer to study
   ○ **(a)** in a study group.
   ◉ **(b)** alone.

22. I am more likely to be considered
   ◉ **(a)** careful about the details of my work.
   ○ **(b)** creative about how to do my work.

23. When I get directions to a new place, I prefer
   ◉ **(a)** a map.
   ○ **(b)** written instructions.

24. I learn
   ◉ **(a)** at a fairly regular pace. If I study hard, I'll "get it."
   ○ **(b)** in fits and starts. I'll be totally confused and then suddenly it all "clicks."

25. I would rather first
   ◉ **(a)** try things out.
   ○ **(b)** think about how I'm going to do it.

26. When I am reading for enjoyment, I like writers to
   ○ **(a)** clearly say what they mean.
   ◉ **(b)** say things in creative, interesting ways.

27. When I see a diagram or sketch in class, I am most likely to remember
   ◉ **(a)** the picture.
   ○ **(b)** what the instructor said about it.

28. When considering a body of information, I am more likely to
   ○ **(a)** focus on details and miss the big picture.
   ◉ **(b)** try to understand the big picture before getting into the details.

29. I more easily remember
   ◉ **(a)** something I have done.
   ○ **(b)** something I have thought a lot about.

30. When I have to perform a task, I prefer to

(a) master one way of doing it.
(b) come up with new ways of doing it.

31. When someone is showing me data, I prefer
(a) charts or graphs.
(b) text summarizing the results.

32. When writing a paper, I am more likely to
(a) work on (think about or write) the beginning of the paper and progress forward.
(b) work on (think about or write) different parts of the paper and then order them.

33. When I have to work on a group project, I first want to
(a) have "group brainstorming" where everyone contributes ideas.
(b) brainstorm individually and then come together as a group to compare ideas.

34. I consider it higher praise to call someone
(a) sensible.
(b) imaginative.

35. When I meet people at a party, I am more likely to remember
(a) what they looked like.
(b) what they said about themselves.

36. When I am learning a new subject, I prefer to
(a) stay focused on that subject, learning as much about it as I can.
(b) try to make connections between that subject and related subjects.

37. I am more likely to be considered
(a) outgoing.
(b) reserved.

38. I prefer courses that emphasize
(a) concrete material (facts, data).
(b) abstract material (concepts, theories).

39. For entertainment, I would rather
(a) watch television.
(b) read a book.

40. Some teachers start their lectures with an outline of what they will cover. Such outlines are
(a) somewhat helpful to me.
(b) very helpful to me.

41. The idea of doing homework in groups, with one grade for the entire group,
(a) appeals to me.
(b) does not appeal to me.

42. When I am doing long calculations,
(a) I tend to repeat all my steps and check my work carefully.

(b) I find checking my work tiresome and have to force myself to do it.

43. I tend to picture places I have been
    - (a) easily and fairly accurately.
    - (b) with difficulty and without much detail.

44. When solving problems in a group, I would be more likely to
    - (a) think of the steps in the solution process.
    - (b) think of possible consequences or applications of the solution in a wide range of areas.

When you have completed filling out the above form please click on the Submit button below. Your results will be returned to you. If you are not satisified with your answers above please click on Reset to clear the form.

Submit Reset

Dr. Richard Felder, felder@ncsu.edu

Learning Styles Results

```
Results for: Ebru Özpolat

ACT                                   X                            REF
      11   9    7    5    3    1    1    3    5    7    9    11
                                <--  -->

SEN                      X                                         INT
      11   9    7    5    3    1    1    3    5    7    9    11
                                <--  -->

VIS   X                                                            VRB
      11   9    7    5    3    1    1    3    5    7    9    11
                                <--  -->

SEQ                                        X                       GLO
      11   9    7    5    3    1    1    3    5    7    9    11
                                <--  -->
```

- If your score on a scale is 13, you are fairly well balanced on the two dimensions of that scale.
- If your score on a scale is 57, you have a moderate preference for one dimension of the scale and will learn more easily in a teaching environment which favors th at dimension.
- If your score on a scale is 911, you have a very strong preference for one dimension of the scale. You may have real difficulty learning in an environment which d oes not support that preference.

We suggest you print this page, so that when you look at the explanations of the different scales you will have a record of your individual preferences.

For explanations of the scales and the implications of your preferences, click on Learning Style Descriptions.

For more information about learning styles or to take the test again, click on Learning Style Page.

APPENDIX C

Table C Whole Learner Profile Table

Learning Resource Type	**TheArtOf PresentingContent**	**RequiredDegreeOf Collaboration**	**TypeOf Coverage**
exercise	sequentially,step-by-step	individual work	practical-known methods,applicable to real world
exercise	sequentially,step-by-step	individual work	theoretical-concepts,discover relationships
exercise	sequentially,step-by-step	team work	practical-known methods,applicable to real world
exercise	sequentially,step-by-step	team work	theoretical-concepts,discover relationships
exercise	sequentially,step-by-step	with the assistance of the teacher	practical-known methods,applicable to real world
exercise	sequentially,step-by-step	with the assistance of the teacher	theoretical-concepts,discover relationships
exercise	globally-with the big picture of subject	individual work	practical-known methods,applicable to real world
exercise	globally-with the big picture of subject	individual work	theoretical-concepts,discover relationships
exercise	globally-with the big picture of subject	team work	practical-known methods,applicable to real world
exercise	globally-with the big picture of subject	team work	theoretical-concepts,discover relationships
exercise	globally-with the big picture of subject	with the assistance of the teacher	practical-known methods,applicable to real world

Table C (continued)

exercise	globally-with the big picture of subject	with the assistance of the teacher	theoretical-concepts,discover relationships
simulation-demo	sequentially,step-by-step	individual work	practical-known methods,applicable to real world
simulation-demo	sequentially,step-by-step	individual work	theoretical-concepts,discover relationships
simulation-demo	sequentially,step-by-step	team work	practical-known methods,applicable to real world
simulation-demo	sequentially,step-by-step	team work	theoretical-concepts,discover relationships
simulation-demo	sequentially,step-by-step	with the assistance of the teacher	practical-known methods,applicable to real world
simulation-demo	sequentially,step-by-step	with the assistance of the teacher	theoretical-concepts,discover relationships
simulation-demo	globally-with the big picture of subject	individual work	practical-known methods,applicable to real world
simulation-demo	globally-with the big picture of subject	individual work	theoretical-concepts,discover relationships
simulation-demo	globally-with the big picture of subject	team work	practical-known methods,applicable to real world
simulation-demo	globally-with the big picture of subject	team work	theoretical-concepts,discover relationships
simulation-demo	globally-with the big picture of subject	with the assistance of the teacher	practical-known methods,applicable to real world
simulation-demo	globally-with the big picture of subject	with the assistance of the teacher	theoretical-concepts,discover relationships
visual resources-figure/diagram/charts (like graphs, tables)	sequentially,step-by-step	individual work	practical-known methods,applicable to real world

Table C (continued)

visual resources-figure/diagram/charts (like graphs, tables)	sequentially,step-by-step	individual work	theoretical-concepts,discover relationships
visual resources-figure/diagram/charts (like graphs, tables)	sequentially,step-by-step	team work	practical-known methods,applicable to real world
visual resources-figure/diagram/charts (like graphs, tables)	sequentially,step-by-step	team work	theoretical-concepts,discover relationships
visual resources-figure/diagram/charts (like graphs, tables)	sequentially,step-by-step	with the assistance of the teacher	practical-known methods,applicable to real world
visual resources-figure/diagram/charts (like graphs, tables)	sequentially,step-by-step	with the assistance of the teacher	theoretical-concepts,discover relationships
visual resources-figure/diagram/charts (like graphs, tables)	globally-with the big picture of subject	individual work	practical-known methods,applicable to real world
visual resources-figure/diagram/charts (like graphs, tables)	globally-with the big picture of subject	individual work	theoretical-concepts,discover relationships

Table C (continued)

visual resources-figure/diagram/charts (like graphs, tables)	globally-with the big picture of subject	team work	practical-known methods,applicable to real world
visual resources-figure/diagram/charts (like graphs, tables)	globally-with the big picture of subject	team work	theoretical-concepts,discover relationships
visual resources-figure/diagram/charts (like graphs, tables)	globally-with the big picture of subject	with the assistance of the teacher	practical-known methods,applicable to real world
visual resources-figure/diagram/charts (like graphs, tables)	globally-with the big picture of subject	with the assistance of the teacher	theoretical-concepts,discover relationships
experiment	sequentially,step-by-step	individual work	practical-known methods,applicable to real world
experiment	sequentially,step-by-step	individual work	theoretical-concepts,discover relationships
experiment	sequentially,step-by-step	team work	practical-known methods,applicable to real world
experiment	sequentially,step-by-step	team work	theoretical-concepts,discover relationships
experiment	sequentially,step-by-step	with the assistance of the teacher	practical-known methods,applicable to real world
experiment	sequentially,step-by-step	with the assistance of the teacher	theoretical-concepts,discover relationships

Table C (continued)

experiment	globally-with the big picture of subject	individual work	practical-known methods,applicable to real world
experiment	globally-with the big picture of subject	individual work	theoretical-concepts,discover relationships
experiment	globally-with the big picture of subject	team work	practical-known methods,applicable to real world
experiment	globally-with the big picture of subject	team work	theoretical-concepts,discover relationships
experiment	globally-with the big picture of subject	with the assistance of the teacher	practical-known methods,applicable to real world
experiment	globally-with the big picture of subject	with the assistance of the teacher	theoretical-concepts,discover relationships
written text	sequentially,step-by-step	individual work	practical-known methods,applicable to real world
written text	sequentially,step-by-step	individual work	theoretical-concepts,discover relationships
written text	sequentially,step-by-step	team work	practical-known methods,applicable to real world
written text	sequentially,step-by-step	team work	theoretical-concepts,discover relationships
written text	sequentially,step-by-step	with the assistance of the teacher	practical-known methods,applicable to real world
written text	sequentially,step-by-step	with the assistance of the teacher	theoretical-concepts,discover relationships
written text	globally-with the big picture of subject	individual work	practical-known methods,applicable to real world
written text	globally-with the big picture of subject	individual work	theoretical-concepts,discover relationships
written text	globally-with the big picture of subject	team work	practical-known methods,applicable to real world

Table C (continued)

written text	globally-with the big picture of subject	team work	theoretical-concepts,discover relationships
written text	globally-with the big picture of subject	with the assistance of the teacher	practical-known methods,applicable to real world
written text	globally-with the big picture of subject	with the assistance of the teacher	theoretical-concepts,discover relationships
lecture	sequentially,step-by-step	individual work	practical-known methods,applicable to real world
lecture	sequentially,step-by-step	individual work	theoretical-concepts,discover relationships
lecture	sequentially,step-by-step	team work	practical-known methods,applicable to real world
lecture	sequentially,step-by-step	team work	theoretical-concepts,discover relationships
lecture	sequentially,step-by-step	with the assistance of the teacher	practical-known methods,applicable to real world
lecture	sequentially,step-by-step	with the assistance of the teacher	theoretical-concepts,discover relationships
lecture	globally-with the big picture of subject	individual work	practical-known methods,applicable to real world
lecture	globally-with the big picture of subject	individual work	theoretical-concepts,discover relationships
lecture	globally-with the big picture of subject	team work	practical-known methods,applicable to real world
lecture	globally-with the big picture of subject	team work	theoretical-concepts,discover relationships
lecture	globally-with the big picture of subject	with the assistance of the teacher	practical-known methods,applicable to real world
lecture	globally-with the big picture of subject	with the assistance of the teacher	theoretical-concepts,discover relationships

CURRICULUM VITAE

PERSONAL INFORMATION

Surname, Name: Özpolat, Ebru
Nationality: Turkish (TC)
email: eozpolat@gmail.com

PUBLICATIONS

1. Özpolat E. and Akar, G. "Automatic detection of learning styles for an e-learning system", Computers & Education, In Press, Corrected Proof, Available online 27 March 2009
2. Özpolat E. and Akar, G. "Constraint determination algorithm for query reformulation for personalized educational relational databases," IV International Conference on Multimedia, Information and Communication Technologies in Education, Seville, Spain, Nov. 2006