INDIVIDUALIZED SELECTION OF LEARNING OBJECTS

A Thesis Submitted to the College of

Graduate Studies and Research

In Partial Fulfillment of the Requirements

For the Degree of Master of Science

In the Department of Computer Science

University of Saskatchewan

Saskatoon

By

JIAN LIU

Permission to Use

In presenting this thesis in partial fulfilment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.

Requests for permission to copy or to make other use of material in this thesis in whole or part should be addressed to:

Head of the Department of Computer Science

University of Saskatchewan

Saskatoon, Saskatchewan S7N 5C9

Abstract

Rapidly evolving Internet and web technologies and international efforts on standardization of learning object metadata enable learners in a web-based educational system ubiquitous access to multiple learning resources. It is becoming more necessary and possible to provide individualized help with selecting learning materials to make the most suitable choice among many alternatives.

A framework for individualized learning object selection, called Eliminating and Optimized Selection (EOS), is presented in this thesis. This framework contains a suggestion for extending learning object metadata specifications and presents an approach to selecting a short list of suitable learning objects appropriate for an individual learner in a particular learning context. The key features of the EOS approach are to evaluate the suitability of a learning object in its situated context and to refine the evaluation by using available historical usage information about the learning object.

A Learning Preference Survey was conducted to discover and determine the relationships between the importance of learning object attributes and learner characteristics. Two weight models, a Bayesian Network Weight Model and a Naïve Bayes Model, were derived from the data collected in the survey. Given a particular learner, both of these models provide a set of personal weights for learning object features required by the individualized learning object selection.

The optimized selection approach was demonstrated and verified using simulated selections. Seventy simulated learning objects were evaluated for three simulated learners within simulated learning contexts. Both the Bayesian Network Weight Model and the Naïve Bayes Model were used in the selection of simulated learning objects. The results produced by the two algorithms were compared, and the two algorithms highly correlated each other in the domain where the testing was conducted.

A Learning Object Selection Study was performed to validate the learning object selection algorithms against human experts. By comparing machine selection and human experts' selection, we found out that the agreement between machine selection and human experts' selection is higher than agreement among the human experts alone.

Acknowledgements

I would like to take this opportunity to express my deep gratitude to my supervisor Dr. Jim Greer not only for his guidance, knowledge, stimulation, and support but also for his understanding, patience, and encouragement. Without his effort I would not be able to finish this project.

I am grateful to my thesis committee members, Dr. Julita Vassileva, Dr. Michael Horsch, and Dr. Kalyani Premkumar for their valuable suggestions and comments. Their time and efforts are greatly appreciated. Special thanks to Dr. Horsch for the inspiring conversations and discussions with him.

Thank you to current and former members of the ARIES lab as well as staff members of the Department of Computer Science for providing a supportive, resourceful, and friendly environment, especially Jan Thompson, Christopher Brooks, Diego Zapata, Collene Hansen, Lori Kettel, and Gina Koehn.

I would like to thank the undergraduate students in the Department of Computer Science who participated in the learning preference survey, as well as the graduate students and staff members at the University of Saskatchewan who completed the validation study. Special thanks to Fuqu Wu for her support with the survey delivery and data collection.

Thanks to Dr. Permanand Mohan for helpful suggestions and comments he made during his visit at the University of Saskatchewan.

The National Sciences and Engineering Research Council of Canada (NSERC) granted the Postgraduate Scholarship to me for my graduate study. Their financial support is greatly appreciated.

My special friends, Gail and Peter Fehr, Yibin Su, Wenling Chen, and Jeanette Dean, encouraged me all the way through the journey of completing this study. Thank you. Your interests in my work and faith in my capability meant a lot to me.

Finally, I must thank my family, especially my patient husband and my understanding son, for their unconditional support and love.

Table of Contents

List of Figures

List of Tables

List of Acronyms

ADL	Advanced Distributed Learning
ARIADNE	Alliance of Remote Instructional Authoring and Distribution Networks for Europe
AEH	Adaptive Educational Hypermedia
CanCore	Canadian Core Learning Resource Metadata Application Profile
EOS	Eliminating and Optimized Selection
IEEE	Institute of Electrical and Electronics Engineers
IMS	IMS Global Learning Consortium, Inc. Instructional Management Systems
ITS	Intelligent Tutoring System
ITSC	Learning Technology Standards Committee
LOM	Learning Object Metadata
RDF	Resource Description Framework
RS	Recommender System
SCORM	Sharable Content Object Reference Model
XML	Extensible Markup Language

Chapter 1

Introduction

Rapidly evolving internet and web technologies have unlocked tremendous possibilities in the world. The movement towards web-based education is significant among them. Through the internet, digital educational materials can be delivered by online learning systems effectively and affordably to a learner almost anywhere and at any time. Because of their convenience and flexibility, online learning systems have been increasingly gaining attention from both education providers and consumers.

Online digital learning resources are commonly referred to as learning objects in E-Learning community. They offer a new way of thinking about learning content. Actually, learning objects can be educational components presented in any format. Learning objects are commonly stored in learning object repositories which facilitate various functions, such as learning object creation, submission, search, comment, review, etc. Several learning object repositories are accessible in both subscription and open-source forms. The data model that is used to describe learning objects is called learning object metadata. Metadata is an important characteristic of any learning object repository since it facilitates the search for relevant learning objects. Most current learning object repositories assume that content searches are performed by a human teacher or a learner and are not well designed for fully automatic computer-based retrieval.

A world-wide effort has been made in developing learning object metadata standards and specifications. The focus of learning object metadata standardization is to improve reusability and interoperability of learning objects. Learning objects that comply with these standards and specifications can be easily discovered, acquired, and reutilized. This enables the sharing and exchange of learning objects across different learning systems and also provides learners access to multiple learning resources.

As a result of such ubiquitous access, learners in an online virtual course may have more diverse backgrounds than those in a traditional course. Different learners have their distinctive characteristics and learning styles. Their learning goals, knowledge level, preferences, and desired level of academic achievement may not be the same. The resources individuals may have (bandwidth, software, hardware) can also vary. The expected benefit of a learning object and the learning effect gained from it are usually different from learner to learner. The traditional one-for-all approach to content selection becomes inadequate in an online learning environment. Because of the limitation of time and capability, however, it is almost impossible for a learner (or a teacher) to go through all available learning materials to find the most suitable learning objects. Selecting the most suitable learning objects among all candidates for individual learners becomes imperative in an online learning environment.

1.1 Suitability of Learning Objects

The suitability of a learning object has various manifestations, such as its appropriateness with respect to the learning goal, its usefulness and helpfulness for learners, pedagogical value, general popularity among learners, and endorsement by teachers. It requires a comprehensive understanding of the learning object and the learning context in which it may apply. The context here refers to the learner and the environment in which he/she resides or currently operates.

A suitable learning object should be able to stimulate learners' motivation. In other words, a learning object might be considered less suitable for a learner if it makes his/her learning procedure more difficult or less interesting. This effect can be seen through the content selection and sequencing of learning objects. Let's consider the task of learning the "loop structure" in JavaScript. For a learner who has some programming background, explanations about iteration structures and related terminologies are usually not necessary and may be excessively verbose and boring. However, for a learner with no coding experience, a detailed explanation should be very helpful. Without explanation the novice learner might feel confused and become frustrated. A student who has difficulty in learning and is willing to work very hard to master the concept might prefer a learning object containing many examples that illustrate various formats for applying iteration structures and explore the difference between them; while a learner who does not want to spend much time might be annoyed by reading many similar examples.

The way in which the content of a learning object is presented may have a significant effect on motivating learning [7] [9] [11] [35]. An appropriate presentation style of a learning object may make the learning easer or more interesting. If an elementary school student or a marginally educated person gets a tutorial that is written in a way that demands a high reading level, they might give up their learning attempt simply because they find the concept is too hard to understand. An easy-to-read document may make the learning much easier. Depending on their learning and cognitive style, some learners may prefer precise elaborative text descriptions; some would rather watch a video; some love to play around with simulations; and some feel nothing is better than diagrams or flow charts. A learning object that meets a learner's preference will stimulate his/her learning interests and might be more suitable for this learner.

The main purpose of a learning object is to teach a specific domain concept. Its pedagogical value is a very important feature. Not only should a more suitable learning object deliver the content that meets the learner's learning goal, but also it should be high in quality. More often the latter has a dominating effect. For example, a learning object developed by an author who has insufficient domain knowledge may be less valuable than a highly appraised learning object designed by an expert.

To sum up, a perfectly suitable learning object for a particular learner should possess the following features:

- It presents the knowledge that the learner wants to learn;

- It can be effectively and efficiently delivered in the learner's environment. i.e. it is affordable in the sense of finance and time to the learner, and it can be presented on the platform the learner has;
- It is appropriate to the learner's knowledge level, which includes domain knowledge and reading capability, etc.;
- Its presentation style matches the learner's preferences to the greatest extent possible;
- It has high pedagogical value.

Unfortunately, such an ideal learning object can rarely be found in the real world. Usually a learning object has only some of those desired features. Moreover, some features of a learning object contribute positively to its suitability, while others contribute negatively. Let us consider the following situation: A student likes video clips. His reading level in English is very high, and his listening level in English is low. If we have a video clip with extensive voice and little content display versus a detailed text document, and both of them are in English, which one is more suitable for this learner? Perhaps, the student is able to learn better by reading the text based material than by watching the video clip. In this situation, the learner's language skill has a stronger affect on learning outcomes, so that the needs associated with this aspect should have more weight.

In a more complicated situation, a learning object whose features apparently match a learner's preferences might not be the best choice for the learner. For example, we try to make a selection for a learner who prefers to watch videos more than to read text documents. A video clip evaluated negatively by other similar learners would not be considered as a good choice, while a text recommended by the instructor might be a better one.

In addition, the suitability of a learning object may change when it migrates to a different context. An excellent learning object can become totally useless in a different context, thus it is not suitable at all. For example, a learning object written in Chinese cannot make any sense for a learner who does not understand Chinese. An interesting simulation program designed for Windows machines can be helpless on a Linux machine. A well designed video clip is not profitable for a learner who cannot afford time to download it. A vivid animation of DNA replication won't do any good for a learner who wants to learn how to build a personal web page.

Individualized learning object selection is a complicated and difficult procedure. Many factors have to be taken into account. It is not enough to find the best match between the features of learning objects and the requirements of the current context. Constraints such as learning objectives must always have higher priority. Besides, a feature of a learning object might have a different effect on learning for different learners. In order to select the most suitable learning object for a specific learner in a given learning situation, we have to be able to identify important features that have stronger effects on the learning and try to accomplish the requirements associated with these features.

1.2 Research Goals

This thesis research aims at designing and developing a practical approach for dynamically selecting the most suitable learning objects for a given context in a web-based educational system. The following goals will be addressed:

- Extend existing learning object metadata specifications to meet the requirements of individualized learning object selection.
- Provide an approach to the selecting of a short list of suitable learning objects appropriate for the learner and the learning context.
- Implement a learning object selector and then verify and validate the approach by comparing its behaviour against human experts' judgment.

1.3 Organization of Thesis

Chapter 2 provides background about learning objects and learning object metadata, as well as adaptive systems such as Intelligent Tutoring System (ITS), Adaptive Educational Hypermedia (AEH), and Recommender System (RS). Chapter 3 proposes a framework for individualized learning object selection, which includes initial considerations on extending learning object metadata specifications and the Eliminating and Optimized Selection (EOS) approach. A Learning Preference Survey is discussed in Chapter 4. The survey data is analyzed, and Bayesian weight models for implementing individualized selection are elicited. Chapter 5 presents the results and analysis of the verification and validation experiments. Chapter 6 outlines the conclusion of this thesis research.

Chapter 2

Background

2.1 Learning Objects and Learning Object Metadata

2.1.1 Learning Objects

"Learning object" is the term that is widely used to refer to educational materials. The Learning Technology Standards Committee (ITSC) of the Institute of Electrical and Electronics Engineers (IEEE) defines a learning object as any entity, digital or non-digital, that may be used for learning, education or training [16]. According to this board and vague definition, almost everything could be considered a learning object. A traditional text book, a web page, a piece of multimedia content, a software tool and even a person, an event, or a place can all be considered learning objects. The IEEE definition has been highly criticized. It fails to become an authentic and universally accepted definition. Consequently, various definitions, which narrow down the scope, have been created by different groups of practitioners [12] [39] [45]. Wiley proposes a working general definition of a learning object – "any digital resource that can be reused to support learning" [45]. He believes that his definition is sufficiently narrow because it captures the critical attributes of a learning object, *reusable*, *digital*, *resource*, and *learning*, and complies with the IEEE definition as well.

Most definitions agree that learning objects are reusable digital resources. However, the multifarious definitions focus on different aspects and suggest diverse structures of learning objects due to assorted origins of researchers and practitioners and of their distinct interests. Some people, who probably have backgrounds related to computer science, like to draw parallels between learning objects (educational materials) and computer science concepts. They concentrate on the functionality of learning objects and their interactions with each other or with environment. For example, Robson proposes viewing learning objects in an object-oriented model [37]. They have methods to function or act. Downes suggests that learning objects can be thought as "small, self-reliant computer programs" and can interact with the Learning Management System [10]. In this community, *modularity*, *interoperability*, and *discoverability* of learning objects are considered as important attributes [13].

Instead of the technical aspects, the structure of learning objects, i.e. components of learning objects are emphasized in some definitions. The value of a learning object depends on the learning that a learner can gain from it [22]. Merrill suggests that a knowledge objects consists of an entity and its parts, properties, kinds (classes), associated activities, and associated processes[27] [28]. L'Allier elaborates NETg's learning object structure in [22]. NETg's philosophy is that a learning object will teach the intended skill and provide verification that learning has taken place by using valid assessments. This philosophy is reflected by three elements of learning objects:

- ***Objective***: describes the intended criterion-based result of a learning activity;
- ***Learning Activity***: teaches towards the objective;
- ***Assessment***: determines if the objective has been met.

The goal of the thesis research is to develop an approach for individualized learning object selection. A learning object has to be evaluated to decide its suitability. NETg's learning object structure appears rational and essential for this purpose. In the scope of this research, we assume that all learning objects consist of those three elements as specified in Netg's definition.

2.1.2 Standards and Specifications about Learning Objects

As countless learning objects are available around the world, their reusability, interoperability, and portability become critical and beneficial. To address this issue, international efforts have been made on developing standards and specifications about learning objects since late 1990's. IEEE Learning Technology Standards Committee, IMS Global Learning Consortium, Inc., and CanCore Initiative are organizations active in this area.

IEEE LOM Standard is a multipart standard, which is composed of Standard for Learning Object Metadata Data Model, Standard for XML Binding and Standard for RDF Binding. The first part of the standard, IEEE 1484.12.1 LOM Data Model standard [16], has been accredited and released. The LOM Data Model is the core of existing metadata specifications. It defines a hierarchical structure for describing a learning object. In a LOM instance, relevant characteristics of learning object are represented by data elements that are grouped into nine categories. *Figure 2-1* depicts the overall structure of LOM Data Model.

The metadata specification developed by IMS and ARIADNE was the origin of IEEE LOM Standard. Since then, IMS has released various versions of IMS specification based on updates of IEEE LOM Standard development. Besides IMS Learning Resource Meta-Data Information Model (IMS Metadata Specification) [18], current IMS specification includes documents defining other useful operations such as learning content packaging and simple sequencing.

The IMS Content Packaging Specification [17] provides the functionality to describe and organize learning materials. A Content Package refers to a unit of reusable educational content, which is a logical directory (tree structure) that consists of a special XML file describing the content organization and resources in a Package, as well as associated physical files. The IMS Simple Sequencing Specification [20] defines a method for arranging the order of learning materials. Learning content in Simple Sequencing is also organized into a hierarchical structure. Alternative learning materials are siblings in the tree. Whether a piece of content is selected or skipped and when it is delivered depend on a set of predefined rules (conditions). The main concern of these two specifications is still interoperability. IMS content package and sequence representation may be interchanged between compliant systems.

The IEEE LOM standard and IMS specification are both complex and general. There are many possibilities left open for interpretation. CanCore addresses this issue with its synthesis efforts that include guidelines for selecting elements, refinements of

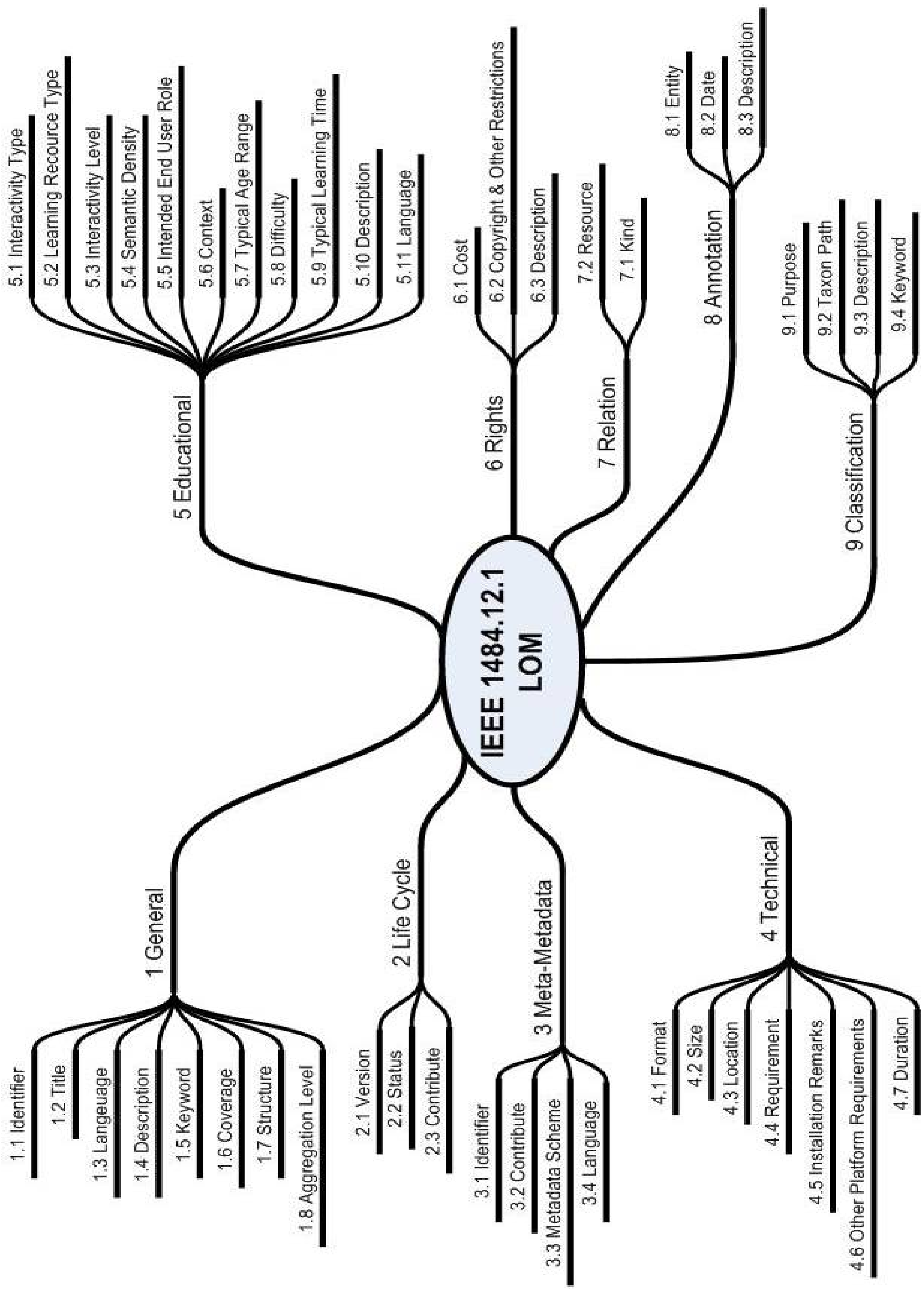


Figure 2-1 Overview of LOM Structure (Reprinted from [19])

definitions, examples, technical implementation notes, and vocabulary recommendations [14]. CanCore is an instantiation of the LOM standard that occupies the middle ground between this standard and the concrete work for building interoperable metadata records.

The Advanced Distributed Learning (ADL) Initiative is another organization working with IEEE and IMS closely. While CanCore focuses on semantics and interpretation, ADL puts efforts on technical issues. ADL's Sharable Content Object Reference Model (SCORM) bundles or integrates a collection of specifications and standards into a collection of "technical books", a set of interrelated technical standards, specifications and guidelines designed to meet high-level requirements for learning content and systems [1]. It is often illustrated as a bookshelf holding nearly all of the specifications come from other organizations including IEEE, IMS, etc. The SCROM consists of three main topics, Content Aggregation (CAM), Run-time Environment (RTE), and Sequencing and Navigation (SN). The technology developments from those groups are integrated within a single reference model to specify consistent implementations, and additional detail and implementation guidance have been added.

Existing standards and specifications about learning objects focus on facilitating search, evaluation, acquisition, and reuse of learning objects such that they can be shared and exchanged across different learning systems. From a pedagogical point of view, however, it falls short in several important areas [29]. The pedagogical information available in the standard is very limited. Some important educational characteristics such as *Pedagogical Objective* and *Prerequisite* are not included. Some attributes are ambiguous or inadequate. For example, definitions for Interactivity Level, Semantic Density, and Difficulty are too ambiguous to keep consistent for different content authorisers, and using *Keyword* to carry information for learning object discovery may lead to incorrect results.

In addition to the inadequacy in terms of information requirements for educational design, we notice that there is one thing that is neglected in existing standards and specifications: comparing all available learning objects and selecting the most suitable one(s) in a given context. Having the Content Packaging Specification and the Simple Sequencing Specification, the delivery of learning content for different learners can vary to a certain degree. Individualization, however, is difficult to be achieved yet. The IMS content package and sequence representation are all predefined static structures. Dynamically changing instruction based on learners' status and availability of learning material is not captured.

Some research involving individualization in web-based educational systems has been conducted. Recker et al. propose to use "non-authoritative" data elements, which can be defined differently for the variety of review areas of learning object, to capture the context of use [34]. These "non-authoritative" data elements may reduce the interoperability and reusability that current standards and specifications are after. McCalla and Brooks argue that the metadata cannot capture enough information and that it is impossible to keep perfect consistency between content and corresponding metadata [3] [26]. They suggest that more information about content such as users' characteristics and interaction with the content should be accumulated and attached to the content. Because of the promise of exchanging and sharing learning objects, however, this standardized metadata approach is well accepted around the world. To meet the requirements of individualized learning object selection, extending existing standards and specifications to include more information such as contextual requirements and historical usage would be one direction worthy to explore.

2.2 Adaptive Systems

Individualization is the goal feature of assorted adaptive systems. Intelligent Tutoring Systems (ITS), Adaptive Educational Hypermedia (AEH), and Recommender Systems (RS) are noticeable examples. Various technologies and techniques have been developed to support services that accommodate different needs of users.

2.2.1 Intelligent Tutoring Systems and Adaptive Educational Hypermedia

Instead of the traditional one-for-all teaching model, Intelligent Tutoring Systems (ITSs) provide individualized instruction or tutoring to learners. The key component for achieving this functionality is the learner model. The information about a learner, such as knowledge in a certain domain, learning style, and relevant personal characteristics, is kept in the learner model and used by the system to identify the particular needs of the learner. Combined with applying pedagogical principles, suitable learning materials or activities are selected, and then they are organized and delivered in an appropriate or preferred way to the learner [15] [43].

Adaptive Educational Hypermedia (AEH) originates from the combination of ITS and hypermedia presentations of learning content. Along with the rapid growth of the World Wide Web, the advantages of hypermedia have been clearly realized and research in AEH has become increasingly popular [4] [5] [6]. Like an Intelligent Tutoring System, Adaptive Educational Hypermedia provides individualized service, adaptive presentation and navigation, to a learner based on the learner's characteristics captured in the system.

The common ground of these two types of system is reflected not only by their capability of adaptiveness, but also by their limitation:

- The adaptation can be achieved only among the local alternatives.
- Rules and conditions for learning resource selection and organization are predefined.
- The decision made in the system mainly relies on the built in virtual expert.

Non-local authorised learning materials and activities, therefore, are difficult to be integrated dynamically in such systems.

2.2.2 Recommender Systems

Recommender Systems (RSs) have been revolutionizing the way shoppers and information seekers find what they want. Recently recommender technology is being deployed in more and more online business entities to best articulate and accommodate customers' tastes. According to the techniques applied, they can be divided into three major categories: content-based, collaborative, and hybrid recommendation [2] [36].

- Content-based recommendation is derived from Information Retrieval. This type of systems identifies and extracts features of items and builds matching model for them. User profiles including information about their preferences, tastes are collected as well. Recommendations are made based on comparison of user's preference and item's features.

- Systems that use collaborative filtering techniques are also called clique-based systems. The main idea of collaborative filtering is grouping like-minded users together. It is assumed that users who had similar choices before will make same selection in the future. Collaborative recommender systems give users suggestion by observing the neighbour of the user.
- Due to the nature of the techniques deployed, the content-based recommender systems have obvious limitations. This type of recommender systems doesn't perform well if the content of items cannot be easily extracted. Collaborative recommender systems also face some challenges. One is the well-known cold-start problem, the situation that there is not enough users' feedback about the item. This type of system will perform poorly when having an unusual user because it will be difficult to find neighbours for the user. Hybrid recommendation mechanisms attempt to deal with some of these issues and overcome drawbacks of pure content-base approach and pure collaborative approach by combining the two approaches.

2.2.3 Learning Object Selection

Strictly speaking, items that ITS and AH deal with can be all considered as learning objects, but these are not the type of learning objects that this thesis research focuses on. Learning objects discussed here have the following characteristics: all learning objects are described by their metadata so that they can be easily discovered, requested, and reused across different learning systems. The candidate set for the selection can be very large. Individualized selection or recommendation in this domain, therefore, is more challenging.

A group led by Duval E. in Katholieke Universiteit Leuven, Belgium has been actively working on learning object selection [31] [46]. They use Contextualized Attention Metadata (CAM) to capture information about actions through out learning object lifecycle including creation, labelling, offering, selecting, using, and retaining. Four metrics using LOM and CAM are proposed for ranking and recommendation: Link Analysis Ranking, Similarity Recommendation, Personalized Ranking, and Contextual Recommendation. These metrics calculate various categorized rankings for learning objects, such as popularity ranking, object similarity based on number of downloads, etc. How these different rankings contribute to the learning object selection and how to combine them together are still questions faced by this group.

Researchers and developers in e-learning have begun attempts to apply recommender technologies, especially collaborative filtering, in learning object recommendation. McCalla proposed an enhanced collaborative filtering approach, called the ecological approach, for designing e-learning systems [26]. The key aspects of his approach involve gradually accumulating information and focusing on end users. Recker et al. are developing and evaluating their Internet-accessible system called Altered Vista where collaborative filtering techniques are applied within an ad hoc designed metadata structure [33] [34]. Manouselis et al. performed a case study on data collected from users of European Schoolnet's CELEBRATE portal to determine an appropriate collaborative filtering algorithm [24] [25]. Lemire's group proposed the RACOFI, Rule-Applying Collaborative Filtering, architecture to customize learning object selection

[23]. Their recommendation is narrowed down and personalized by combining the collaborative filtering algorithm with an inference rule system.

In all these collaborative learning object recommendation systems, the key problem of cold start has not been addressed. Some more recent work has been done towards solving this problem. Tang et al. practice collaborative filtering in their evolving research paper recommender system [40] [41]. They emphasize the importance of pedagogical characteristics and try to use artificial learners to overcome the cold-start problem. The domain of their system, however, is limited to research papers thus the factors that influence the paper selection are much less complicated than those affecting learning object selection. Tsai, Wang et al. take the hybrid approach [42] [44]. Similar to collaborative learning object recommendation systems, correlation-based algorithms are used to calculate a helpfulness score via analyzing similar learners' feedback. In addition, preference-based algorithms enhance the selection with learners' preference. A Learner Preference Pattern is kept for each learner to record the preference history, which is generated and updated according to the learner's preference feedback. If a learning object is selected, and positive feedback is given; an increment is made to preference scores of all features of the learning object. The combination of scores of a learning object determined by the two algorithms decides its rank in the recommendation result. Their preference-based algorithm helps with the cold-start problem. However, all features of the selected learning object are treated equally. This might be the cause for more error recommendations that they admit to being generated by their system in some cases.

2.3 Conclusion

Standardized learning object metadata makes sharing and exchange of learning objects possible. Because of the large potential quantity of available candidates, learning objection selection can be more challenging. Ad hoc designs or approaches are no longer feasible in this setting. In addition to traditional techniques for achieving adaptiveness (e.g. user modelling), techniques developed in recommender systems are becoming more explored in the e-learning area. The key issue is what information to store and how to use it. This will be discussed in this thesis research.

Chapter 3

A Framework for Individualized Selection of Learning Objects

3.1 Information Requirements for Individualized Selection

The existing learning object metadata specifications have defined a set of attributes that describe learning objects. The suitability of a learning object, however, is a contextual feature. It can be decided only when the learning object is situated in a certain context. To determine the suitability of a learning object, information about the learner and learning situation is necessary in addition to information about the learning object itself. Besides feature and requirement matching, the suitability of a learning object depends on some features that are more difficult to describe and measure. The historical usage and historical measures of suitability of learning objects can provide valuable information for optimizing selection.

The following three subsections discuss attributes related to the three areas required for individualized selection. It is not necessary to get explicit input for every attribute in order to perform the individualized selection. Some of them can be inferred from other attributes, and also sometimes the selection has to be done while some information is lacking. This kind of information ideally should become part of learning object metadata, and the results of this research will hopefully influence future work on metadata standards.

3.1.1 Information about Context

Learning Objective

The learning objective includes the information about the subject or topic the current learner is going to learn. In formulating this objective there is a negotiation between the learner's preferred topic and that of the curriculum specialist. The preferred objective of the curriculum specialist is based primarily on the knowledge state of the learner and the general learning goal.

Learner Characteristics

The learner is central to the context. Information about the learner plays a significant role in determining the most suitable learning object. Theoretically, the more that is known about a learner, the better the selection that can be made for him/her. However, many criteria and constraints may interfere with the selection, and sometimes situational variables add a great deal of complication to the decision. An analysis of the literature reveals the following learner characteristics:

> ***Learner Type:*** provides information about the learner's category. For example, a learner could be a part/full time university student, a high school student, or salesman. This can be used to infer some other information.

Background: gives information about related knowledge or experiences of the learner (such as the major of a university student, the domain area in which the learner has extensive knowledge, etc.). This information can be useful when comparing learning objects involving background knowledge or skill prerequisites.

Knowledge in Related Area: provides information about the learner's knowledge level in the domain area related to the topic the learner intends to study. For example, if a learner wants to learn iteration structures in JavaScript, information about his/her general experience with programming may be very helpful in learning objective selection.

Details of Domain Knowledge: includes a model of the learner's detailed domain specific knowledge. In the previous example, a model showing the learner's knowledge about the many specific concepts in JavaScript and HTML will be useful.

Preferred Language: contains a list of languages that the learner prefers for learning materials (e.g. English, French, and Chinese).

Reading Level: refers to the learner's capability of understanding written materials with varies level of difficulty, which is associated with the preferred language(s). This may be inferred from other attributes.

Listening Level: refers to the learner's capability of understanding verbal vocal materials with varies level of difficulty, which is associated with the preferred language(s). This may be inferred from other attributes.

Reading Speed: refers to the learner's speed of reading, which is associated with the preferred language(s). Categories of slow, normal, and fast should be sufficient. This can be inferred from other attributes.

Preferred Presentation Style: specifies the learner's preferred way in which the content of a learning object is presented (e.g. text, diagram/picture, video, etc. and their combination).

Learning Style: indicates the way in which the learner learns a new concept or knowledge (e.g. example lover, concept analyst, brief reader, etc.).

Study Attitude: reveals the learner's attitude towards studies (e.g. hard worker, eager learner, interest driven, lazy student, etc).

Academic Achievement Goal: specifies the academic goal the learner wants to achieve, such as exceptional mark, excellent mark, or good mark, pass the class, etc.

General Academic Achievement: gives information about the learner's past academic performance. For example, the learner's grade of courses taken before is mostly exceptional, excellent, good, pass, or fail, etc.

History of Using Learning Objects: includes a list of learning objects that have been accessed.

Resource

Resource describes information about things that may affect the learner's access to technology. Whether resources are consistent with the requirements of a particular learning object will strongly affect its suitability. Related factors could include the following:

Computer Environment: refers to the hardware, software, network access, and other related conditions.

Financial Situation: gives information about the learner's financial restriction. For example, how much the learner can afford to spend on learning goals. If the learner obtains learning materials via an organization, this will refer to how much the organization would spend for this purpose.

Time: provides information about the time the learner wishes to spend on a learning object. A lengthy learning object is probably not a good choice for a learner who has very limited time to devote to learning the concept.

3.1.2 Information about Learning Objects

A number of standards and specifications for learning object metadata have been developed. This standardization effort focuses on promoting reusability and interoperability through defining text-based tags for categorizing and annotating learning objects, which facilitate learning object discovery and exchange across different learning objects repositories sufficiently. To achieve individualized selection, however, extension and modification are required for some attributes. Below are attributes needed for individualized learning object selection.

Pedagogical Objective

The pedagogical objective of a learning object describes the concept that the learning object presents and what is expected to be achieved. Learning objects are generally categorized now but need a more refined categorization into different groups according to their pedagogical objectives. A learning object cannot be suitable if its educational objective does not match the learner's learning objective.

In current existing specifications, pedagogical objectives of learning objects are not well addressed. The educational objective of a learning object might be indirectly inferred from attributes such as *keyword* and *description* if a human teacher is involved. *Description* is difficult to be used for automatic learning object comparison and selection. *Keyword* is not sufficient and sometimes could mislead the learner and result in unexpected learning outcomes. An ontology of pedagogical objectives may serve much better to link learning objects. Much work on ontologies is ongoing so this will improve in the future.

Environment

The environment is about the technical requirements needed for presenting the learning object. For example, the learning object may need some specific hardware and/or software support. If this environment cannot be made available to the learner, the learning object is useless.

In the existing specification, the related information can be determined from attributes *requirement*, *otherPlatformRequirements,* and their sub-entries.

Cost

The cost refers to the price of the learning object. Only the learning objects that the learner can afford are further considered.

This attribute has been included in the existing specifications.

Language

The language in which the content of the learning object is presented is one of decisive factors of the suitability. Only learning objects constructed in the language that the learner can understand become potential candidate for selection.

The existing specifications have this attribute defined.

Expected Reading Level

The expected reading level indicates the reading capability that the learning object requires the learner to have. The reading level affects the learning ability of a learner. Sometimes this influence can be very strong.

In the current existing specifications, the expected reading level is not defined. Instead attributes *context* (the level of education) and *typicalAgeRange* are used. Learners in the same category or in the same age, however, may have different reading ability. Their reading ability actually plays a more important role.

Prerequisite

The prerequisite specifies the knowledge needed by the learning object. For example, a learner who wants to study data structures must have basic knowledge about programming. The gap between the prerequisite of a learning object and a learner's knowledge level may cause frustration.

The prerequisite is not defined in the existing specifications, but it is a very important factor for deciding the suitability of a learning object for a specific learner.

Typical Learning Time

Approximate time needed for working with the learning object. This attribute can be used to decide whether the learning object is suitable. It can also be used to evaluate the effort a learner contributes to the learning object and from this infer the learner's evaluation of the learning object. Usually, a learner would spend more time on a learning material that is found to be useful and interesting.

This attribute has been included in the existing specifications.

Presentation Style

The presentation style describes the way in which the content of the learning object is presented. For example, a learning object can be presented in plain text, figure/diagram, video, slides, and their combination. It can be detailed description, brief outline, example, animation, etc. Each individual may have his/her own preference. A learning object that is presented in the format preferred by the learner is more motivational. This feature can be very important in some domain areas and for a certain type of learners.

In the existing specifications, information about presentation style can be found in entry *learningResourceType*.

3.1.3 Information about Learning Object Usage History

Some features relating to quality and appropriateness of a learning object, which may impact its suitability in the given context, may not be readily describable by an author or evaluator. Much useful information can be indirectly gathered from prior experiences with the learning object by learners and instructors. In some situations such information is of the utmost importance. This kind of information should be recorded in the learning object usage history and attached to the learning object [26]. Some researchers call this type of information situational metadata or attention metadata to record [31] [46].

Previous Learners

In this research work we have decided that the information about previous learners comprises a list of learners who have accessed the learning object in the past. Along with each learner in the list, information about the following aspects is also recorded:

Accessing Time: records the start time and the duration when the learning object is accessed by the learner.

Learner Status: contains snap shots of the learner's state (learner model) before and after accessing the learning object.

Interactions: records actions the learner makes while accessing the learning object, such as help requests, outside references to other resources, as well as the duration that the learner stayed with the learning object.

Evaluation: reveals the learner's opinions about the learning object. This can be direct feedback obtained from the learner or implicit inference from the learner's actions.

Achievement: shows the assessment result of the learner after working with the learning object, such as post quiz mark.

Previous Instructors

For this thesis we have decided that the information about previous instructors contains a list of teachers who have accessed the learning object and their evaluation or endorsements.

Statistics

The final category of learning object usage data we need for this thesis research includes some general usage statistics. The statistics information is accumulated when the learning object is accessed. It can be helpful when more detailed information is not available.

General Popularity: provides information about how often the learning object has been selected from among all comparable candidates for all types of learners.

Categorized Popularities: provides information about how often the learning object has been selected for certain types of learners.

While the information about learning objects, learning contexts and learning situation described in this Section (3.1) seems quite comprehensive, other variables or features could be used to characterize these phenomena. The general framework presented here is based on these variables or features, but could readily be modified or expanded to incorporate other features deemed relevant.

3.2 The Eliminating and Optimized Selecting (EOS) Approach

To determine the suitability of a learning object, we have to evaluate how well its features meet the needs of the learner in the current context. A natural and simple way to do it is to assign a weight to each feature and then sum up weight * value pairs across individual features. As illustrated in the previous section, however, a learning object feature affects the suitability of the learning object differently in different contexts. A very important feature may become a nonentity when the target learner or the environment where the learner resides changes. It is not feasible to define a fixed weight for each feature that applies to all potential learners and all potential contexts. In the EOS approach, the important features are identified by examining the current context and the weights associated with them are modified dynamically.

3.2.1 Two Steps of the EOS Approach

Attributes of a learning object exerts influence on its suitability in two ways.

- The learning object becomes unsuitable and is eliminated from the candidate list if some important attribute of the learning object and some important requirement of the current context do not match;
- The learning object may become a more suitable choice among all candidates if some attribute of the learning object and some requirement of current context do match.

The attributes that play the eliminating role are called eliminating attributes; and the attributes that help in making better selection are called selecting attributes.

To make use of properties of these two categories of attributes, the selection can be divided into two steps. First, eliminate irrelevant learning objects and reduce the domain of selection. Second, evaluate all learning object candidates in the domain and identify the most suitable one. This two step architecture is depicted in *Figure 3-1*.

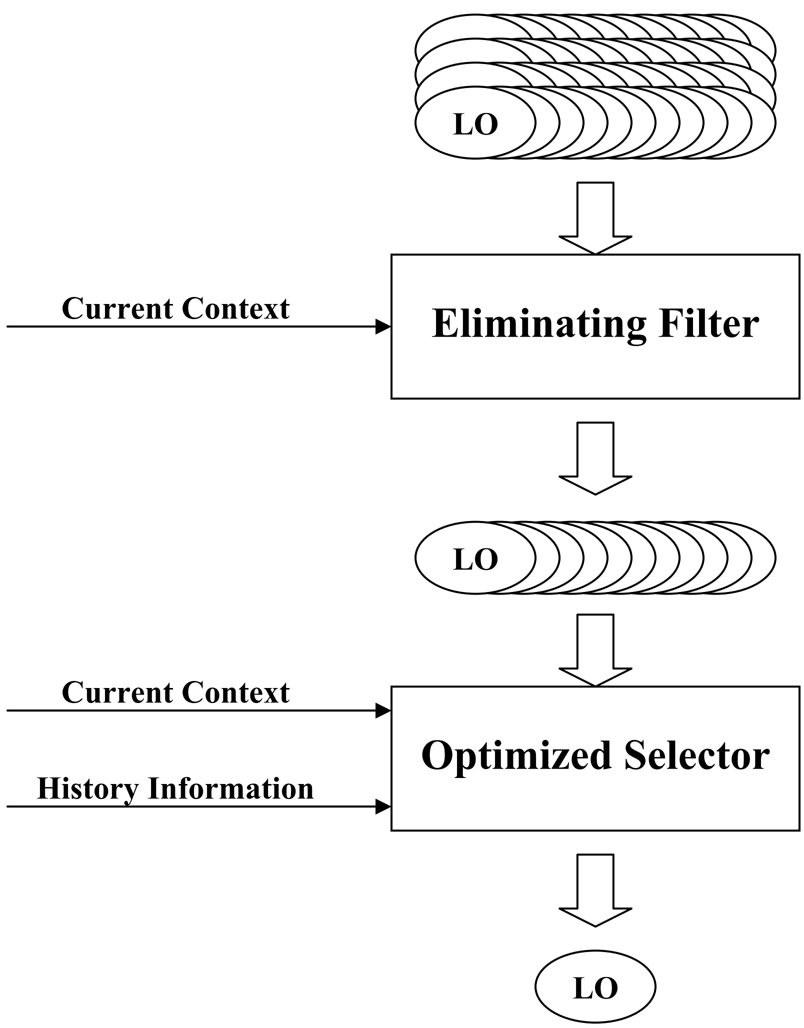


Figure 3-1 Two Steps Structure of EOS Approach

Let e_{final} be the final result of the evaluation, $e_{eliminate}$ and $e_{optimized\text{-}select}$ be the result of the two steps respectively, we have:

$$\mathbf{e}_{final} = \mathbf{e}_{eliminate} \times \mathbf{e}_{optimized-select}$$

3.2.2 Eliminating Irrelevant Learning Object Candidates

Eliminating attributes are constraints and they would normally have only two values: 1 (true) or 0 (false). If the feature of a learning object represented by an attribute satisfies the requirement of the current context, it has value 1 (true), and the learning object will be selected to perform further comparison; otherwise, its value is 0 (false), and the learning object is eliminated. Attributes in this category are used to eliminate irrelevant learning objects from comparison procedure. They could be:

- The pedagogical objective. For example, if a learner wants to learn loops in JavaScript, a learning object about Java Exceptions will not be helpful.
- The language in which the content of the learning object is presented. Obviously, a learning object written in Chinese will not be useful for a learner who can understand only English.
- Hardware, software and other environment condition requirements. If a learning object cannot run on learners' machine, it becomes useless.
- The financial cost of the learning object.

Let $a_{eliminate\ i}$ be the value of an eliminating attribute, result of this step of evaluation ($e_{eliminate}$) is

$$\mathbf{e}_{\text{eliminate}} = \prod_{i} \mathbf{a}_{\text{eliminate i}}$$

where $a_{\text{eliminate}} \in \{0, 1\}$.

When the quantity of available learning objects is limited, some negotiations can be performed in this step to adjust the selection range. For example, increasing the limit of financial cost, or removing some constraints on software by installing necessary software.

The constraints listed above are some simple and typical constraints that we selected to demonstrate our approach in this research. There may be other constraints that can be used for elimination.

3.2.3 Optimized Selecting

Selecting attributes help make the selection among all relevant learning objects. The contribution that each attribute makes to the selection is reflected by its importance, which can be indicated by its weight assignment.

Similarly, Let α_i be the value of a selecting attribute, result of this step of evaluation (e_{select}) can be

$$\mathbf{e}_{\text{select}} = \sum_{\text{i}} \mathbf{w}_{\text{i}} \times \alpha_i$$

where $w_i, \alpha_i \in [0, 1]$.

Among selecting attributes, some of them are much more important than the others. Through examining a known context, especially the learner features, the important attributes might be decided. *Table 3-1* gives some examples. These important features could dominate learning object selection. Therefore, the weights assigned to the important attributes should be much higher.

Table 3-1 Examples of Important Feature Identified from Learner Features

Learner Features	**Important LO Attributes**
Elementary school students	Reading level
Low educated learners	Reading level
Non-first language learners	Easy to read
Learners with weak prior knowledge	Prior knowledge review
Hard workers with low achievement	Enough simple examples
Eager learners	Complete coverage
Hard workers with high achievement	Complete coverage and comprehensive examples
Low motivated learners	Non-lengthy material
Learners working with exercises	Material with examples
Review step learners	Summary

As described in previous section, the final selection of the most suitable learning object may be improved by using information about previous usage of learning objects, such as experts' evaluation, similar learners' experience, and popularities of learning objects. Influences from these aspects can be negative, and they may also be assigned with different weights to distinguish their importance.

Let β_j be the an adjustment value, $e_{optimize}$ be the result of total adjustment, and v_j be the weight assigned to each adjustment, we have

$$e_{optimize} = \sum_j v_j \times \beta_j$$

where v_j, $\beta_j \in [0, 1]$.

Then the result of optimized evaluation ($e_{optimized\text{-}select}$) becomes

$$e_{optimized-select} = e_{select} + e_{optimize}$$

Based on evaluation of all candidates, the most suitable one(s) is recommended.

3.3 Scope of Thesis

Eliminating irrelevant learning objects can be easily combined with searching or query in Learning Management Systems, so we will not make further discussion on this part. The rest of the thesis will focus on optimized selection.

Chapter 4

Eliciting a Model from Users to Implement Individualized Selection

One learning object could be more suitable than others for a specific learner in a certain learning situation. What features of a learning object determine its suitability? How important are these features in a particular context? How is the importance of those features related to the characteristics of learners? A Learning Preference Survey was conducted in order to discover and determine some of these relationships.

The questionnaire for the Learning Preference Survey consisted of three parts with forty-eight questions. The first fourteen questions in part one were about the student's background, including academic achievement and other information related to online learning. Part two consisted of nine questions asking for the student's opinion on the importance of features of learning material, such as presentation format, access speed, etc. The last part of the survey provided two scenarios of online learning. Fourteen statements were made about scenario one, and the students were asked to what extent they agree or disagree with those statements. This part was used for conforming and validating students' answers with the questions answered in part two. The eleven questions associated with scenario two were designed for determining each student's degree of trust in recommendations from different people, for example, teachers' recommendations or peer recommendations. Please refer to *Appendix A* for details of the questionnaire, consent form and ethics committee approval.

4.1 Study Sample

The Learning Preference Survey was conducted online. It was made accessible to all students registered in introductory computer science courses (CMPT100 and CMPT111) at the University of Saskatchewan in the fall of 2004. The reason for making this survey available only to students in their first computer science class was to leave out the influence that the level of the course may have on the selection preference for learning material.

During the survey period, one hundred and three students completed the survey questionnaire. Each signed (digitally) a consent form and was offered an honourarium of $5. The characteristics of this group of students are summarized in *Figure 4-1*. From the Figure, one can see that the distributions on registration status, net access, and first language are extremely skewed, so these variables were ignored in further analysis.

4.2 Importance of Features of Learning Materials

In part two of the Learning Preference Survey, we asked students to rate the importance of eight aspects of features of learning material. To avoid counting careless answers to these questions, each statement was given in a positive and negative form to describe the importance of a feature. Students indicated the degree to which they agreed/disagreed

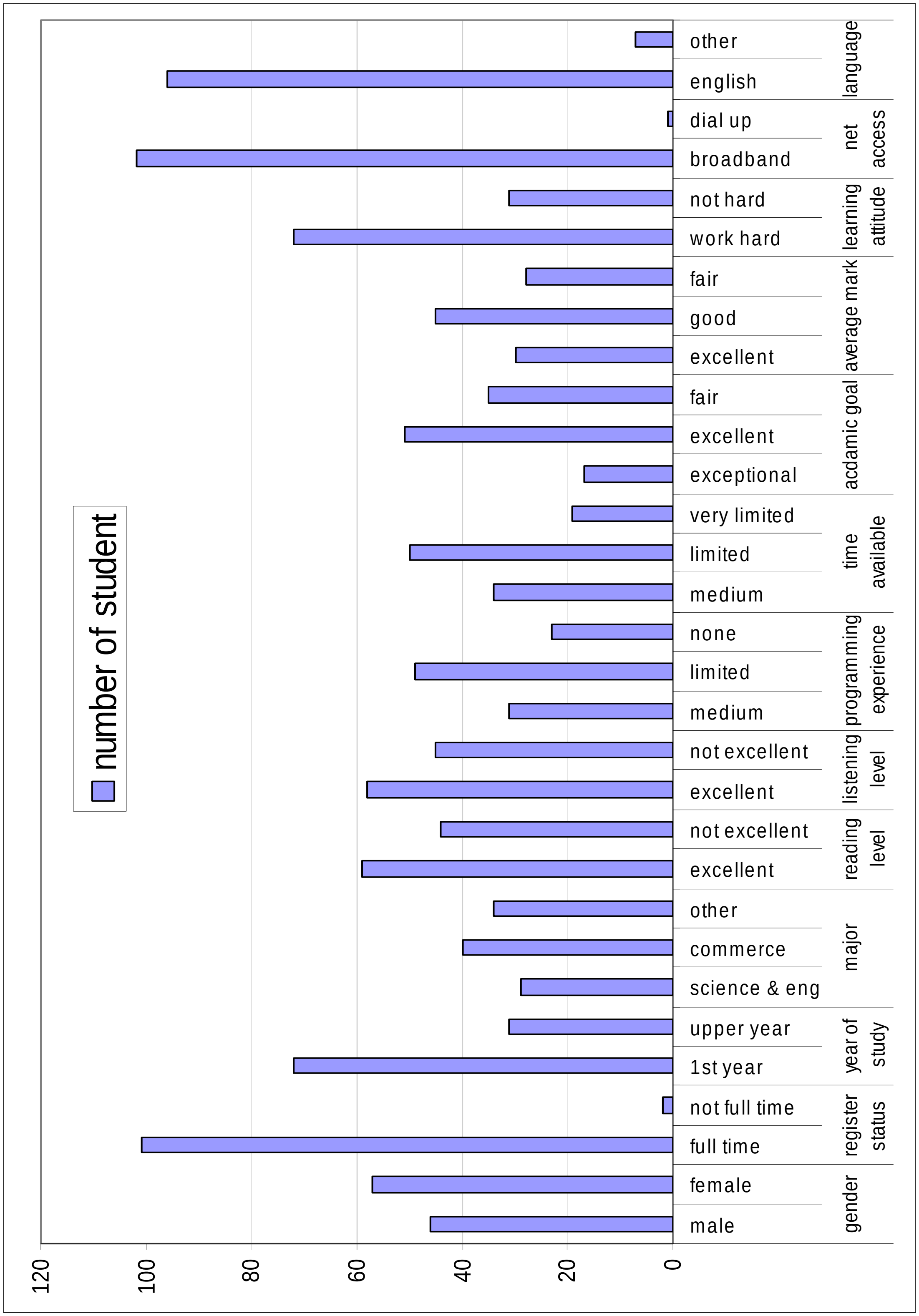


Figure 4-1 Natures of the Population of the Survey

with each statement. If a student's answers to the pair of corresponding items are contradictory, that student's answer was considered invalid. No student selected answers that directly contradicted each other in the survey, and thus there is some assurance that the students answered thoughtfully.

Figure 4-2 shows how important students thought each feature of learning material to be in general. For example, about 30% of the students rated the presentation format of learning material to be very important, about 56% considered it important, and about 14% didn't think it was important at all.

In *Chapter* 3 we pointed out that some features of a learning object may be more important than others for a specific learner in a certain situation. In other words, a learning object feature becomes more or less important along with the change of learner or situation. The survey data supports this idea. *Figure 4-3* and *Figure 4-4* are two examples that illustrate how students' opinion on importance of features of learning materials varies with their programming experience and study major respectively. From *Figure 4-2* we can tell that generally speaking about 30% of students think that the format of learning material is a very important factor. If we group students according to their major, we will find that 46% of science & engineering students consider that format are very important while only 22% of commerce students think format is very important (*Figure 4-4*). Therefore the format is more important to a science & engineering student than a commerce student. Similarly, we can presume that required study time of learning materials is a more important factor for a student with less

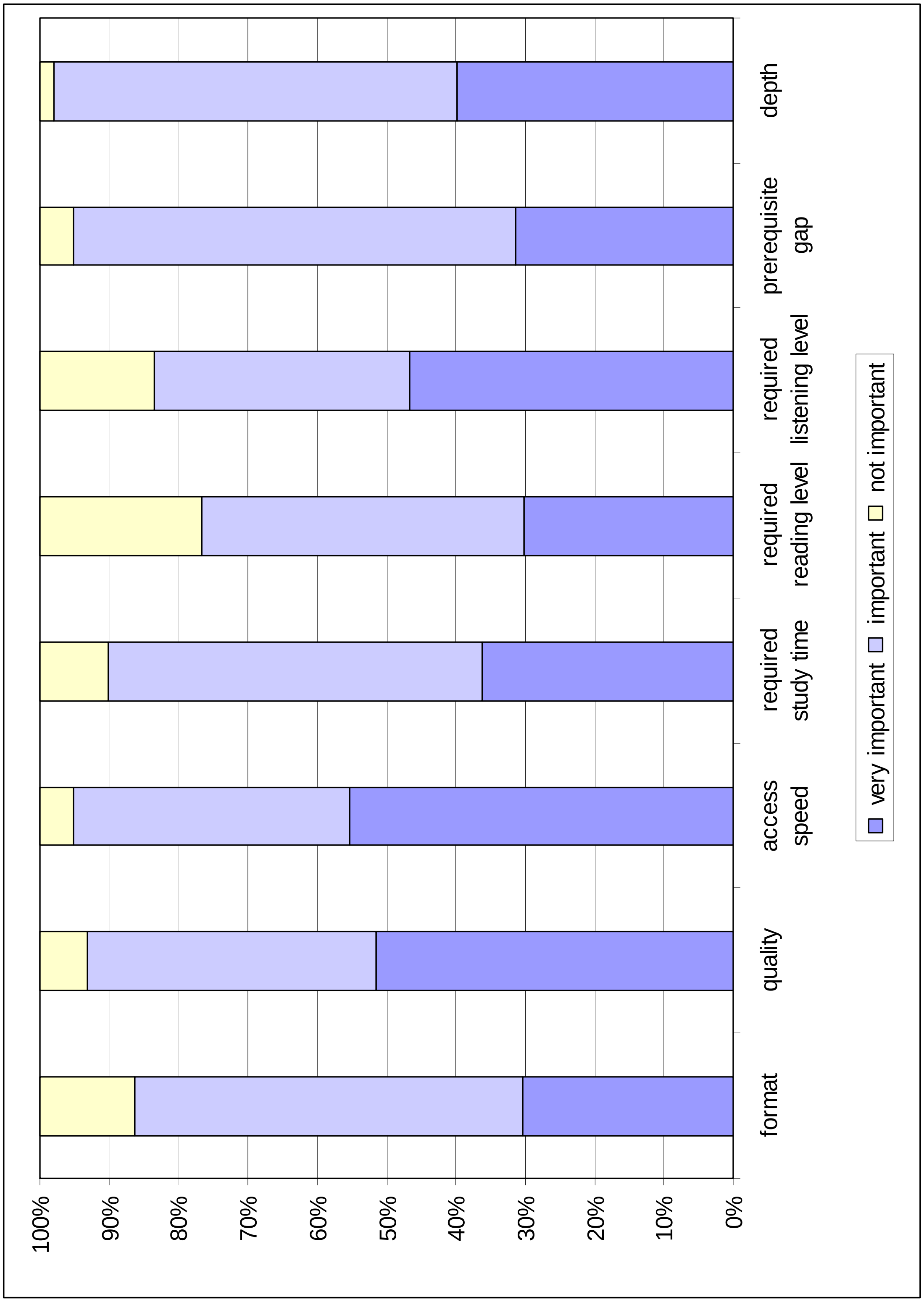


Figure 4-2 Students' General Opinion on Importance of Features of Learning Material

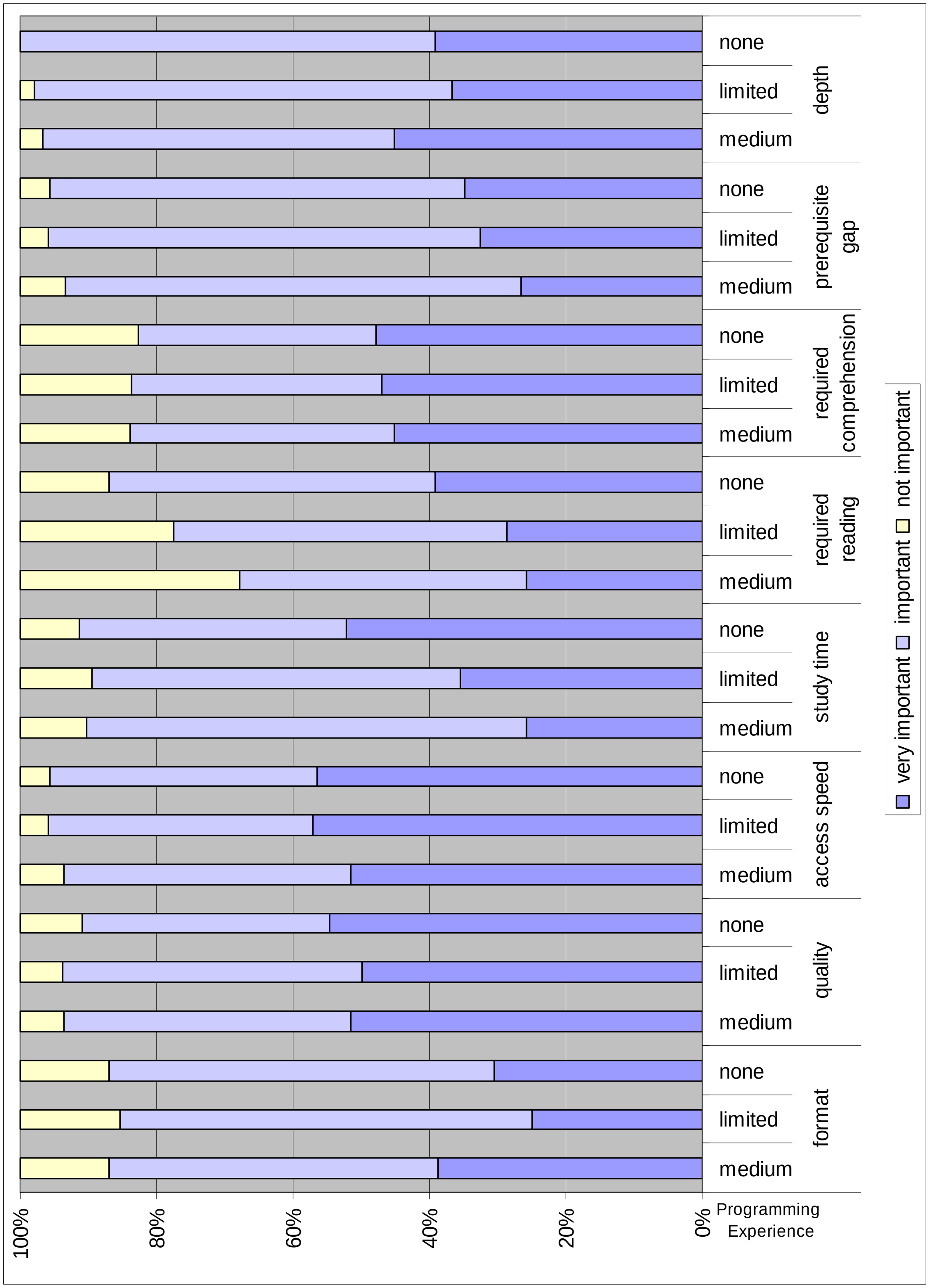


Figure 4-3 Students' Opinion on Importance of Features of Learning Material Varies with Their Programming Experience

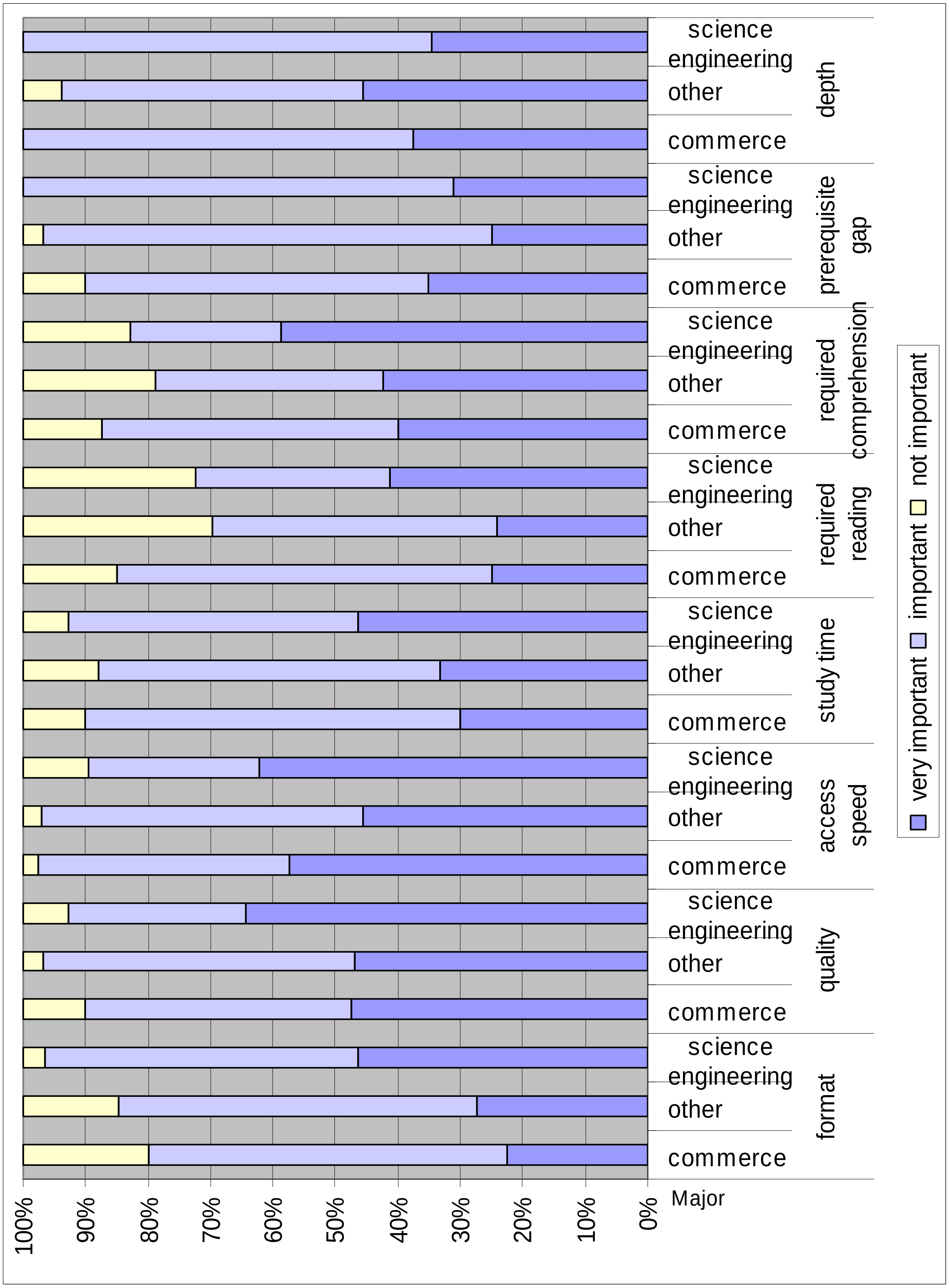


Figure 4-4 Students' Opinion on Importance of Features of Learning Material Varies with Their Major

programming experience because about 52% students who have no programming experience believe that this factor is very important but only about 25% students with moderate programming experience have the same opinion (*Figure 4-3*).

4.3 EOS Weight Models

In the EOS approach, the importance of each feature of a learning object is indicated by its weight assignment. In order to carry out individualized learning object selection, the importance of features of a learning object needs to be decided by examining the learner's characters. This gives rise to the requirement of dealing with uncertainty associated with each individual learner's characteristics and preferences.

In this section, we discuss two models derived from the survey data to carry out this task.

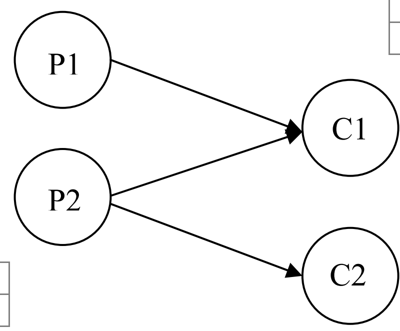
4.3.1 Bayesian Belief Networks

A Bayesian network, also called belief network, is a data structure that represents the probabilistic dependencies among variables. In the network, nodes are variables, arcs specify dependences between variables, and conditional probability tables give numerical expressions of the dependences [8] [21] [32] [38].

In the simple example two-layer Bayesian network shown in *Figure 4-5*, there are four variables, P1, P2, C1, and C2. Both P1 and P2 have direct impact on C1, and C2 is depends only on P2. Node P1 and node P2 are thus called parent node of node C1, and they are both root node of the net because they have no parents. Node C1 and C2 are named leaf node or child node.

p(P1=1)	p(P1=2)	p(P1=3)
0.5	0.1	0.4

p(P2=A)	p(P2=B)
0.3	0.7



P1	P2	p(C1=I)	p(C1=N)
1	A	0.2	0.8
1	B	0.6	0.4
2	A	0.5	0.5
2	B	0.3	0.7
3	A	0.8	0.2
3	B	0.1	0.9

P2	p(C2=I)	p(C2=N)
A	0.7	0.3
B	0.4	0.6

Figure 4-5 A Simple Example of a Bayesian Model

The probability distribution of each variable is illustrated by its conditional probability table (CPT). For example, suppose states of variable P2 could equal A or B, and the states of variable P1 could be 1, 2, or 3 with prior probabilities as shown in the tables in the left side of *Figure 4-5*. Suppose further that both variable C1 and C2 can have value I or N. Complete CPTs are shown for variable C1 and C2 on the right side of the figure. From the CPT associated with variable C1, we know that when P1 = 3 and P2 = A the probability for C1 = I is 0.8.

Before a Bayesian Net can be used, all CPTs must be specified. The information required can be obtained either from domain experts or from empirical data. The initial distribution might not be accurate if the data set is not big enough or if the expert is not experienced enough, but it gradually becomes better as the network is updated with more data [32] [38].

A Naïve Bayes Net is a simple Bayesian Network. It only contains a single root node. All leaf nodes in a Naïve Bayes Net always have one parent. In this case the CPTs are simple to specify and computation is O(n).

4.3.2 Bayesian Weight Model

Our Bayesian Weight Model is a two-layer Bayesian net. There are two types of node in the model, learner characteristics nodes and learning object feature nodes. Learner characteristics nodes, representing aspects of the learner, are root nodes in the net. The states of learner characteristics nodes are the possible values associated with the characteristics, such as *male* and *female* for *gender*. Learning object feature nodes are associated with features of a learning object, and they are leat nodes in the network. States of all learning object feature nodes are specified to be either *very important* or *not very important*. The probability of a feature being very important is used as the weight of this feature.

After eliminating three categories over which the sample from the survey had responded with extremely uneven distribution, ten candidate learner characteristics variables were left: *gender*, *year of study*, *major*, *reading level*, *listening level*, *and programming experience*, *time available*, *academic goal*, *average mark*, and *learning attitude*. All eight learning object features (*format, quality, access speed, required study time, required reading level, required listening level, prerequisite gap, depth*) are candidates for learning object feature nodes.

Table 4-1 Statistic Analysis Results of the Survey

Student Character	Critical Value (p=0.15)	χ^2 Value of Test							
		format	quality	access speed	required study time	required reading level	required listening level	prerequisite gap	depth
gender	2.07	0.76	0.86	1.03	1.23	1.51	0.39	0.23	0.02
study year	2.07	0.79	0.04	0.00	0.62	0.02	0.06	0.55	1.05
major	3.80	**4.70**	2.35	1.90	2.05	2.80	2.61	0.84	0.86
reading level	2.07	0.22	0.21	0.02	0.03	0.01	0.36	0.27	0.04
listening level	2.07	0.36	0.44	0.70	0.19	0.04	0.17	0.14	0.00
programming experience	3.80	1.67	0.13	0.25	**4.00**	1.22	0.04	0.47	0.57
time available	3.80	3.30	0.91	2.13	**4.00**	3.04	**4.58**	3.59	2.47
academic goal	3.80	2.31	0.09	1.71	**3.80**	0.58	1.44	**7.02**	**10.14**
average mark	3.80	3.75	1.36	2.75	0.90	2.36	**4.00**	0.28	1.82
learning attitude	2.07	0.19	0.23	2.00	**2.74**	**2.17**	0.15	1.37	0.22
Note: Values in bold font are significant at p < 0.15.									

Support from nonparametric statistical tests was drawn to discover the connections between learner characteristics variables and learning object feature variables. For each learner characteristics variable, a chi-square (χ^2) independence test was run over every learning object feature variable to determine whether they are statistically dependent. *Table 4-1* lists the chi-square values of all tests.

To include any variable thought to have some degree of influence, we choose 0.15 as the significance level. The critical values for tests are then determined by combining with the number of degrees of freedom. When the chi-square value of a test over a student character variable and a learning object feature variable exceeds the corresponding critical value, the two variables are considered dependent and a link is set in the Bayesian Weight Model between the nodes representing those variables.

For example, the chi-square value of the test over *major* and *format* is 4.70, which is greater than the critical value 3.80, thus a link is added where *major* becomes a parent node of *format* in the model. If a variable that is not dependent on any variable in another category, such as *gender* or *access speed*, there will be no node in the model for the variable. After drawing links between nodes corresponding to variables with significant relationship, the graph structure of the Bayesian Weight Model (*Figure 4-6*) is built up. We will discuss about where the CPTs come from in *Section 4.3.4*.

When a target learner is decided for the individualized selection, learner attributes in the model are instantiated to a particular state. The probability for each learning object attribute's importance can queried from the model. This probability is assigned as the weight for the learning object attribute. A set of personal weights of learning object features for selection, therefore, is generated for each learner in a particular learning context. These weights are used in the selection sub-step of optimized selection of EOS approach. The selection procedure using this model is referred as Bayesian algorithm.

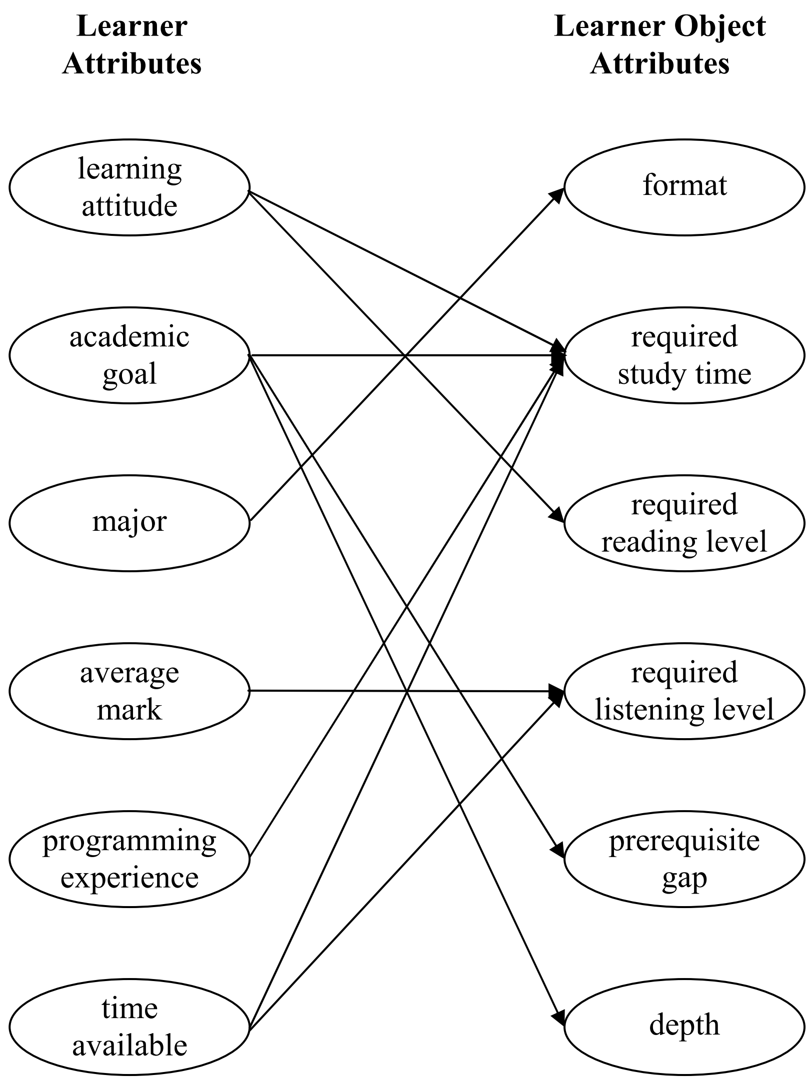


Figure 4-6 Bayesian Weight Model

4.3.3 Naïve Bayes Weight Models

Naïve Bayes weight models are also belief networks. Similar to the Bayesian Weight model, they consist of learner characteristics nodes and learning object feature nodes as well. Instead of identifying individual dependence between learner characteristics variables and learning object feature variables, it is assumed that a learning object feature variable is a function of all learner characteristics variables. Each learning object feature node, therefore, and all learner characteristics form a Naïve Bayes net. The set of Naïve Bayes nets is our weight model. *Figure 4-7* gives two examples of the nets.

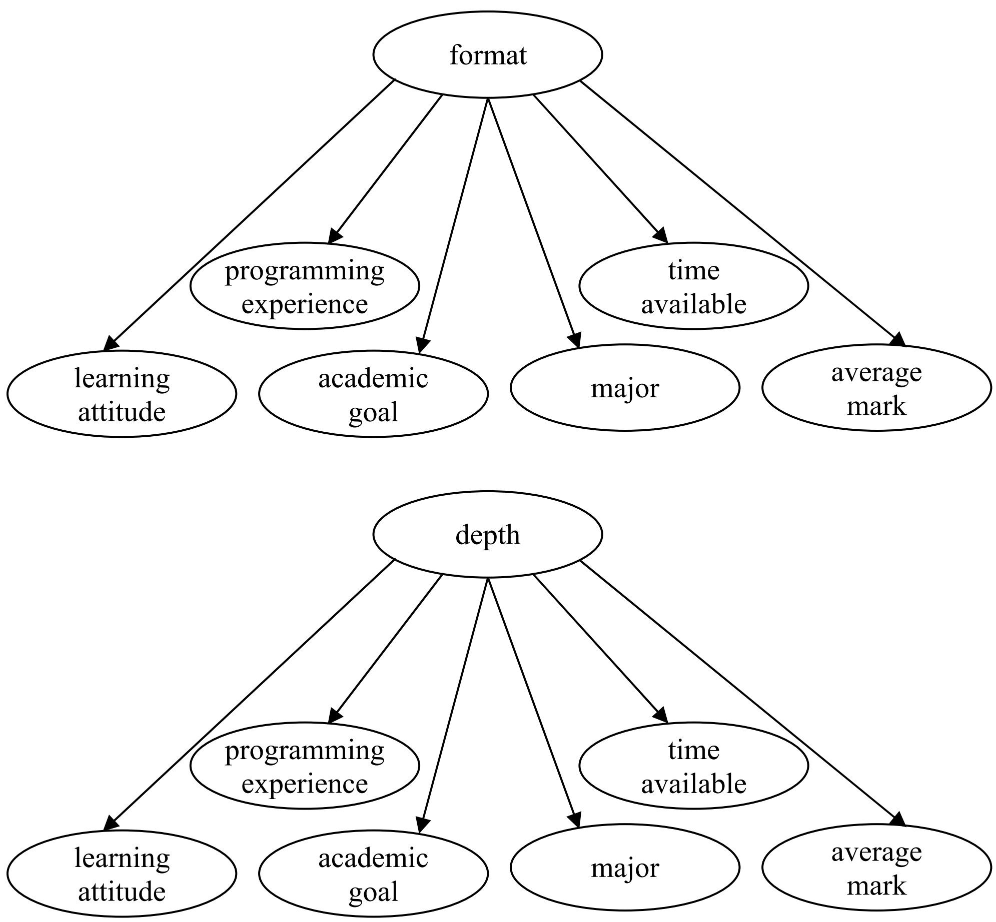


Figure 4-7 Examples of Naïve Bayes Weight Model Nets

Similarly, when we perform the individualized learning object selection for a particular learner, a set of weights can be obtained by querying the Naïve Bayes Weight Model. These weights are also used in the selection sub-step of optimized selection of EOS approach. The selection procedure using this model is called Naïve Bayes algorithm.

4.3.4 Model Implementation

Weight models were constructed in Java using Netica-J, a programmer's library for working with Bayesian networks, developed by Norsys Software Corp[1]. As introduced in *Section* 4.3.1, before the Weight Models can be used, the conditional probability tables in the network must be specified. The initialization of the model should reflect the nature of the domain. This thesis research is focused on first year computer science students, so the survey data is used to initialize our weight models.

Let's look at some nodes in the Bayesian Weight Model to see how the conditional distributions of the model are determined by the survey data. *Major* is a root node in the model, and its distribution over our survey sample is listed in *Table 4-2*. The prior probabilities of each possible value of *major* are then decided and shown in *Table 4-3*.

Table 4-2 Major Distribution

Major	**Science & Engineering**	**Commerce**	**Other**
Number of Students	29	40	34

[1] http://www.norsys.com

Table 4-3 CPT Associated with Node Major

Possible Value	**Science & Engineering**	**Commerce**	**Other**
Probability	0.28	0.39	0.33

The Conditional Probability Tables (CPT) associated with children nodes are derived from the survey data similarly. Node *required listening level*, for example, has two parent nodes, *average mark* and *time available*, and each parent node has three possible values. Therefore there are nine conditioning cases. The possibilities for *Required Listening Level* being very important and less important in each case are calculated from the survey data.

Table 4-4 CPT Associated with Node Required Listening Level

Conditional Case		**Probability**	
Average Mark	**Time Available**	**Very Important**	**Less Important**
excellent	medium	0.70	0.30
excellent	limited	0.46	0.54
excellent	very limited	0.55	0.45
good	medium	0.44	0.56
good	limited	0.30	0.70
good	very limited	0.33	0.67
fair	medium	0.67	0.33
fair	limited	0.41	0.59
fair	very limited	0.80	0.20

There are twelve nodes in the Bayesian Weight Model and forty seven nodes in Naïve Bayes Model. CPTs for all nodes were generated from the data collected from the Learning Preference Survey in the same way described above. Functions in Netica-J library were used to learn CPTs from the survey data automatically.

4.4 Using Historical Information

As discussed in previous chapter, information about previous usage of learning objects, such as experts' evaluation, similar learners' experience, and popularities of learning objects can be used for improving the selection of the most suitable learning object. In the Learning Preference Survey, we also collected students' feedback regarding their trust degree towards various recommendations, which are summarized in *Figure 4-8*.

It is not difficult to tell from *Figure 4-8* that the degree of trust in different recommendations varies. For example, the teacher's recommendation is more trustworthy than a recommendation from a peer student in the same year. To differentiate the influence that various recommendations have on the learning object selection, different weights are assigned to them. Instead of dynamically assigning weights assignment based on the individual learner's characters, static weights are used here. They were calculated as follows:

Weight = (Highly Trust * 2 + Somehow Trust - Somehow Distrust - Highly Distrust * 2) / Sum

The calculation results are listed in *Table 4-5*. Five categories with the highest trust weights were picked. They are *teachers' recommendation*, *recommendation from students with similar academic achievement*, *recommendation from students with similar*

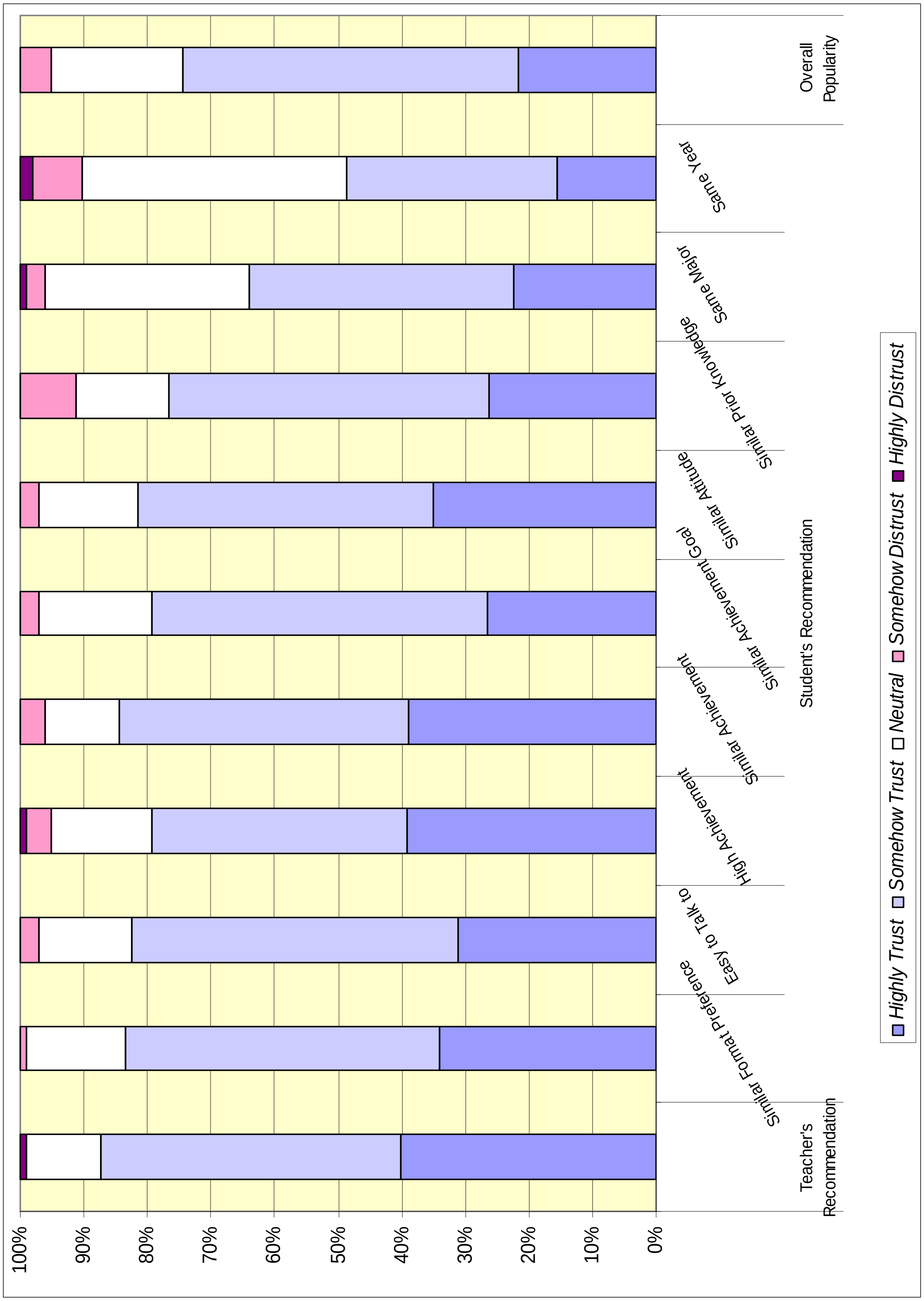


Figure 4-8 Students' Trust Degree to Other's Recommendation

Table 4-5 Weights for Recommendations

Degree of Trust	Teacher's Recommendation (j = 1)	Student's Recommendation				Overall Popularity (j = 6)
		Similar Format Preference (j = 2)	High Achievement (j = 3)	Similar Achievement (j = 4)	Similar Attitude (j = 5)	
Highly Trust	41	35	40	40	36	22
Somehow Trust	48	51	41	47	48	54
Neutral	12	16	16	12	16	21
Somehow Distrust	0	1	4	4	3	5
Highly Distrust	1	0	1	0	0	0
Sum	102	103	102	103	103	102
Weights (v_j)	**1.25**	**1.17**	**1.13**	**1.19**	**1.14**	**0.91**

format preference, *recommendation from students with similar learning attitude*, and *recommendation from students with high academic achievement*. Due to convenience of obtaining *overall popularity*, it was selected as well. That is, six trust categories of recommendation were used in the individualized selection. They play an important role in the adjustment sub-step of optimized selection of EOS approach. This set of statistical weights was used for every learner for whom learning objects were selected.

4.5 Summary

A Learning Preference Survey was conducted to discover and determine the relationships between the importance of learning object attributes and learner characteristics. Two weight models, a two-layer Bayesian network and a set of Naïve Bayes Nets, were derived from the data collected in the survey. Both of them provide a set of personal weights for learning object features required for the selection for a particular learner. Those weights were used in selection sub-step of optimized selection in the EOS approach. The selection procedures using these weight models are referred to as the Bayesian algorithm and the Naïve Bayes algorithm respectively.

Teachers' recommendation, *recommendation from students with similar academic achievement*, *recommendation from students with similar format preference*, *recommendation from students with similar learning attitude*, *recommendation from students with high academic achievement*, and *overall popularity* were selected for the individualized selection. They were used for the adjustment sub-step of optimized selection in the EOS approach.

Chapter 5

Simulated Selection and Validation

Due to lack of uniformity of learning objects that exist and their conformance to the NETg's learning object definition which we adapted, a simulation experiment was performed to test the individualized learning object selection approach. This experiment involved making simulated individualized selections of learning objects for simulated learners. A simulation test-bed was created for this purpose. The purpose of this simulation study was to compare machine selections with human experts' selections. The simulated selection of learning objects and the study will be discussed in this chapter.

As stated in Section 3.3, we focused on optimized selection and set aside the elimination of unusable learning objects by assuming that elimination has been previously applied to the learning object pool. The weight models for the Bayesian networks were derived from the data collected from the Learning Preference Questionnaire conducted with CMPT100 and CMPT111 students. The following assumptions, therefore, were made in the simulated testing:

- All test scenarios are about learning a single concept in introductory JavaScript programming;

- All learning objects are suitable in terms of language, cost, OS, and other eliminating constraints;
- All learning objects available are relevant to the student's learning objective.

5.1 Simulated Test

The simulation test-bed is constructed from simulated learning objects and simulated learners. Both learners and learning objects are represented by collections of metadata. Criteria for generating values for learner and learning object attributes will be described in the following sections.

5.1.1 Learning Object Simulation

According to the previous chapter, six specific learning object features are required for the selection of the most suitable learning object. Their metadata used in generating a simulated learning object contains:

Format: indicates the actual format of a learning object. Its value is randomly picked from its value set: *text, slide, table, diagram, video, audio, simulation, exercise*.

Required listening level: is represented by integer numbers 1 – 5. The bigger the number, the more complex and demanding the listening level. Listening level of a learning object is considered to be a function of its format. A learning object that is not an audio or video will have lower listening requirement.

Required reading level: is similar to the listening level and represented by numbers 1 – 5. It is also related to the format of learning objects. Required reading level of a learning object of type table, diagram, audio, or simulation will be set to a lower range than one with more written content. Learning objects with complex text have a high value for required reading level.

Prerequisite: represents the list of concepts that need to be known as a prerequisite for a learning object. In the experiment, we simplified simulation and evaluation of this attribute. We did not simulate the prerequisite list of concepts required by a learning object and the prerequisite set of concepts mastered by a learner separately. Instead attribute is characterized by the percentage of prerequisites satisfied, which indicates the portion of prerequisites mastered by the learner and has a value between 0% and 100%.

Depth: is the level of difficulty of a learning object and is indicated by an integer number 1 – 5. The bigger the number, the more difficult the learning object.

Required study time: denotes the time in minutes needed to study the learning material. Values for this attribute are allowed to vary between 1 min and 30 min.

Besides these above six learning object features required for the weight calculation in our simulations, six features of historical information (*teachers' recommendation*, *overall popularity*, *recommendation from students with similar academic achievement*, *recommendation from students with similar format preference*, *recommendation from students with similar learning attitude*, *recommendation from students with high academic achievement*) are needed as well. The values for those variables are percentage scores between 1% and 100%.

In addition, *days in use* of a learning object is used to reflect the freshness of the learning object. There is some advantage given to newer learning objects through this parameter, otherwise new learning objects would never be chosen having no usage metadata. It is an integer number picked between 1 and 720.

A simulated learning object is simply a group of attribute that describe its feature and historical usage information. A value was randomly selected for each learning object attribute separately from its value range defined above. Seventy learning objects were created, and *Table 5-1* and *Table 5-2* give some examples. A complete list of all learning objects generated for the study is in *Appendix B*.

Table 5-1 Simulated Learning Objects – Selecting Attributes

ID	Format	Satisfied Prerequisite Portion	Required Reading Level	Required Listening Level	Depth	Required Study Time
LO 00	simulation	0.6192	2	2	5	17
LO 01	video	0.2842	5	4	2	25
LO 02	audio	0.1079	2	2	3	26
LO 03	slide	0.5339	4	2	5	0
LO 04	video	0.7181	2	5	3	11
LO 05	simulation	0.5078	2	2	4	8
LO 06	simulation	0.0724	3	2	2	16
LO 07	text	0.3613	3	3	4	9
LO 08	table	0.4184	2	1	5	19
LO 09	table	0.6945	3	1	3	9
LO 10	video	0.9873	4	4	4	3

Note: Only eleven of seventy generated learning objects are shown here. All seventy are presented in *Appendix B*.

Table 5-2 Simulated Learning Objects – Historical Information

ID	Days in use	Overall Popularity	Recommendation				
			Teacher	Similar Achieve-ment	Similar Format	Similar Attitude	High Achieve-ment
LO 00	478	0.6634	0.5971	0.1416	0.5570	0.8480	0.8558
LO 01	102	0.1416	0.6634	0.5570	0.8480	0.8558	0.0141
LO 02	401	0.5570	0.1416	0.8480	0.8558	0.0141	0.3538
LO 03	611	0.8480	0.5570	0.8558	0.0141	0.3538	0.2626
LO 04	616	0.8558	0.8480	0.0141	0.3538	0.2626	0.5204
LO 05	10	0.0141	0.8558	0.3538	0.2626	0.5204	0.2941
LO 06	255	0.3538	0.0141	0.2626	0.5204	0.2941	0.6192
LO 07	189	0.2626	0.3538	0.5204	0.2941	0.6192	0.2842
LO 08	375	0.5204	0.2626	0.2941	0.6192	0.2842	0.1079
LO 09	212	0.2941	0.5204	0.6192	0.2842	0.1079	0.5339
LO 10	443	0.6149	0.2898	0.2799	0.1036	0.5296	0.7138

Note: Only eleven of seventy generated learning objects are shown here. All seventy are presented in *Appendix B*.

5.1.2 Learning Context Simulation

Based on requirements of optimized selection, learning context is composed of two categories of variables: learner characteristic variables (*Attitude*, *Major*, *Goal*, *Achievement*, *Programming experience*, and *Time available*) and variables for learning object evaluation (*Preferred format*, *Preferred depth*, *learner's listening level*, *Reading level*, *Mastered prerequisite*).

Attitude: refers to learner's learning attitude. It is simplified to a binary score of hard-working or not-hard-working.

Major: is categorized into three groups – science-engineering, commerce, and other.

Goal: indicates learner's academic motivation to be successful – exceptional, excellent, and fair.

Achievement: reflects the learner's grade to date – excellent, good, and fair.

Programming experience: reflects the learner's programming experience. Its value could be medium, limited, or none.

Time available: indicates how busy the learner is. Values for this attribute are medium, limited, or very limited.

Format: represents learner's preferred format of learning materials. Its value set is the same as the value set of learning object format - text, slide, table, diagram, video, audio, simulation, exercise.

Listening level: is denoted by an integer number between 1 – 5. The bigger the number, the higher the learner's listening capability.

Reading level: is similar to the listening level and represented by an integer number between 1 – 5 as well with 5 meaning the student can listen and understand very well..

Mastered Prerequisite: describes the prerequisite set that the learner masters. As explained in *Section 5.1.1*, the simulation and evaluation of prerequisite was simplified. This attribute was not simulated for learners.

Depth: represents the learner's preferred depth of learning materials. It is indicated by integer number 1 – 5. The bigger the number is, the higher the difficulty and the greater the depth level is.

Similar to the learning object simulation, a value was randomly selected for each learner character variable and each variable required for individualized learning object evaluation separately from its value range defined above. A group of these attributes represents a learning context. We have created twenty learning scenarios that could be used for the study. *Table 5-3* shows some of them. The complete list can be found in *Appendix C*.

5.1.3 Simulated Selection

As described in *Section 3.2.3*, the optimized selection step of EOS approach consists of two parts, selection and optimization. Each step results in a score for a learning object. The sum of the score from the both steps becomes the final score for the learning object.

In the selection sub-step, each relevant learning object is evaluated and assigned a score (e_{select}) calculated by the following formula:

$$\mathbf{e}_{\text{select}} = \sum_{\text{i}} \mathbf{w}_{\text{i}} \times \alpha_i$$

Table 5-3 Simulated Learning Contexts

ID	Attitude	Achievement Goal	Major	General Achievement	Programming Experience	Time Available	Reading Level	Listening Level	Preferred Format	Preferred Depth
LC 00	hard	fair	sci & eng	good	medium	limited	5	5	table	3
LC 01	not hard	excellent	commerce	excellent	none	very limited	5	5	diagram	4
LC 03	not hard	fair	sci & eng	fair	medium	very limited	1	2	table	4
LC 15	hard	exceptional	commerce	good	medium	medium	3	4	diagram	4
LC 16	not hard	excellent	sci & eng	good	medium	limited	3	3	diagram	4
LC 17	hard	fair	commerce	good	limited	medium	3	5	simulation	5
LC 18	hard	fair	sci & eng	good	none	medium	3	3	exercise	2
LC 19	hard	excellent	sci & eng	excellent	none	limited	1	4	text	4

In the formula, α_i is a decimal number between 0 and 1 representing degree of match of each selecting attribute of a learning object with a learner's attributes. Evaluation criteria for generating values of α_i are listed in *Table 5-4*. w_i is the weight associated to each selecting attribute for a specific learner, which is obtained by querying weight models (either the Bayesian Model or the Naïve Bayes Model) presented in *Section* 4.3 of this thesis. As described in *Chapter 4*, the weight models were derived from the data collected in Learning Preference Survey.

Table 5-4 Matching Evaluation Criteria (α_i)

Attribute	Evaluation Criteria
format	If a learning object's format is the same as a learner's preferred format, the evaluation value is 1.0; otherwise the value is 0.1.
listening level	If a learner's listening level is not lower than a learning object's required listening level, the evaluation value is 1.0; otherwise, the value is 0.1.
reading level	Same as listing level.
depth	If difference between a learner's preferred depth level and a learning object's depth level is 1, then the evaluation value is 1.0; if the difference is 2, then the evaluation value is 0.5; otherwise, the value is 0.1.
study time	If a learner has very limited study time and a learning object requires more than 20 minutes to study, the evaluation value is 0.1; if a learner has very limited study time and a learning object requires 10 - 20 minutes to study, the evaluation value is 0.5; if a learner has limited study time and a learning object requires more than 20 minutes to study, the evaluation value is 0.5; otherwise, the evaluation value is 1.0.
prerequisite gap	As explained in *Section 5.1.1* and *Section 5.1.2*, the simulation and evaluation of prerequisite was simplified. Satisfied prerequisite percentage simulated for learning objects is actually used for this purpose.

When a learner is targeted, his/her attributes are entered into either the Bayesian Model or the Naïve Bayes Model as findings. The probability of each learning object feature being important in this condition becomes the weight of the corresponding learning object feature for the target learner. This type of weight is a personal weight, and it is different from learner to learner. *Table 5-5* gives characteristics of some learners and weight assignments of learning object features for these learners.

Table 5-5 Weights for Some Learners

Category	**Attributes**	**Learning Context**		
		LC 00	**LC 01**	**LC 03**
Learner Character	Attitude	hard	not hard	not hard
	Achievement Goal	fair	excellent	fair
	Major	sci & eng	commerce	sci & eng
	General Achievement	good	excellent	fair
	Programming Experience	medium	none	medium
	Time Available	limited	very limited	very limited
Weights of Learning Object Feature	Format	0.45	0.23	0.45
	Listening level	0.29	0.50	0.67
	Reading level	0.35	0.22	0.22
	Prerequisite	0.28	0.32	0.28
	Depth	0.39	0.36	0.38
	Study time	0.14	0.33	0.33

The weighted sum of evaluation of all specific learning object attributes required for the selection becomes the score of selection step.

The optimization step decides an adjustment value ($e_{optimize}$) for a learning object based on its usage history using the following formula.

$$e_{optimize} = \sum_j v_j \times \beta_j$$

As discussed in *Section 4.4*, six categories of recommendation (*teachers' recommendation*, *overall popularity*, *recommendation from students with similar academic achievement*, *recommendation from students with similar format preference*, *recommendation from students with similar learning attitude*, *recommendation from students with high academic achievement*) were chosen based on the Learning Preference Survey to optimize the learning object selection. β_j in the formula is a statistic value for each category, which were generated in the simulation for each learning object. Weights (v_j) associated with each recommendation category come from statistical results of the Learning Preference Survey are listed in *Table 4-5*. They are population weights and will not vary for different learners.

In addition to this optimization, it is necessary to introduce a term to sometimes select "newer" learning objects and thus to vary the recommendation. Freshness of learning objects (*days in system*) are taken into consideration to achieve this. The weight assigned to this factor is – 0.001.

Linear combination of all those historical features forms an adjustment value for selection. The sum of evaluation score of a learning object (e_{select}) and its historical adjustment value ($e_{optimize}$) becomes the final score of the learning object. The learning objects with higher score are considered more suitable to the particular learner.

We picked three simulated learners/learning contexts from the twenty simulated learning contexts to perform an individualized selection experiment. For every selected learner, all seventy simulated learning objects were evaluated via the optimized selection step elaborated above. In the experiment, every learning object was evaluated twice for a particular learner using both the Bayesian Weight Model and the Naïve Bayes Model respectively in the selection sub-step to query the personal weights.

The experiment generated six score lists of learning objects, one per learner per algorithm. Ten learning objects that have highest score in each list are considered most suitable learning objects in that situation. The experimental results are summarized in *Table 5-6*. The table is sorted in descending order of scores calculated by Bayesian algorithm for the learning context LC 00. The ten most suitable learning objects for that learning context in each list are shaded. It is not difficult to find out from the data in the table that rank of learning objects varies as learning context changes. It is also interesting to notice that the results produced by the two algorithms for the same learner agreed each other very well (this will be further discussed in *Section 5.3.1*)

Table 5-6 Results of Simulated Individualized Learning Object Selection

Learning Object ID	Learning Context 00		Learning Context 01		Learning Context 03	
	Bayesian Model (score/rank)	**Naïve Model** (score/rank)	**Bayesian Model** (score/rank)	**Naïve Model** (score/rank)	**Bayesian Model** (score/rank)	**Naïve Model** (score/rank)
LO 41	2.66 / 01	2.61 / 01	0.84 / 23	0.76 / 25	1.70 / 06	1.67 / 06
LO 53	2.65 / 02	2.61 / 01	1.35 / 12	1.41 / 11	0.12 / 36	-0.31 / 46
LO 09	2.57 / 03	2.50 / 03	0.81 / 27	0.75 / 26	1.72 / 05	1.38 / 11
LO 11	2.02 / 04	1.95 / 04	-0.03 / 40	0.06 / 37	-0.80 / 52	-1.34 / 55
LO 63	1.89 / 05	1.57 / 08	-1.26 / 54	-1.19 / 53	0.85 / 15	0.57 / 25
LO 69	1.76 / 06	1.72 / 05	0.63 / 29	0.77 / 23	1.68 / 07	1.21 / 12
LO 31	1.59 / 07	1.68 / 06	2.77 / 02	2.68 / 02	1.76 / 04	1.88 / 02
LO 34	1.56 / 08	1.63 / 07	-0.69 / 48	-0.68 / 48	0.79 / 18	0.62 / 24
LO 46	1.35 / 09	1.43 / 09	1.09 / 15	1.13 / 14	1.09 / 13	0.75 / 18
LO 32	1.28 / 10	1.38 / 10	1.36 / 11	1.28 / 12	2.30 / 01	2.76 / 01
LO 67	1.18 / 11	1.25 / 11	-0.04 / 41	0.00 / 40	1.62 / 08	1.56 / 07
LO 61	1.14 / 12	1.08 / 13	0.82 / 26	0.77 / 23	-0.54 / 48	-0.63 / 48
LO 13	1.09 / 13	1.04 / 14	-1.89 / 62	-1.97 / 64	0.26 / 31	0.26 / 34
LO 16	1.07 / 14	1.17 / 12	2.29 / 03	2.20 / 03	-0.37 / 45	-0.32 / 47
LO 62	0.91 / 15	0.98 / 15	1.61 / 08	1.62 / 08	0.25 / 32	0.12 / 37
LO 43	0.85 / 16	0.92 / 16	1.90 / 07	2.07 / 06	-0.97 / 53	-1.02 / 53
LO 27	0.71 / 17	0.78 / 17	2.14 / 05	2.19 / 04	0.52 / 24	0.19 / 36
LO 48	0.71 / 18	0.56 / 21	1.24 / 13	1.14 / 13	1.31 / 10	1.69 / 05
LO 19	0.70 / 19	0.66 / 20	-3.59 / 70	-3.48 / 70	0.14 / 35	-0.14 / 44
LO 28	0.65 / 20	0.36 / 29	-1.94 / 63	-1.85 / 61	0.11 / 38	-0.03 / 41
LO 49	0.62 / 21	0.71 / 18	0.60 / 31	0.65 / 30	1.41 / 09	1.45 / 10
LO 64	0.59 / 22	0.44 / 25	0.63 / 29	0.53 / 31	1.10/ 12	1.46 / 09
LO 17	0.58 / 23	0.41 / 26	0.85 / 21	0.74 / 27	-0.34 / 44	0.09 / 39
LO 00	0.57 / 24	0.67 / 19	3.09 / 01	3.00 / 01	1.77 / 03	1.85 / 03
LO 01	0.56 / 25	0.41 / 26	-1.43 / 58	-1.43 / 57	-1.32 / 59	-1.27 / 54
LO 33	0.56 / 26	0.39 / 28	0.85 / 21	0.74 / 27	0.82 / 17	1.11 / 13
LO 37	0.42 / 27	0.48 / 22	1.08 / 16	1.12 / 15	0.11 / 38	-0.25 / 45
LO 10	0.38 / 28	0.45 / 23	1.97 / 06	1.90 / 07	-1.15 / 56	-1.45 / 59
LO 30	0.37 / 29	0.45 / 23	0.51 / 32	0.45 / 33	0.84 / 16	0.72 / 19
LO 26	0.35 / 30	0.29 / 31	-0.71 / 49	-0.58 / 46	0.42 / 27	0.08 / 40
LO 51	0.31 / 31	0.15 / 34	-0.57 / 46	-0.31 / 44	-1.57 / 64	-1.52 / 60
LO 04	0.24 / 32	0.32 / 30	1.19 / 14	1.10 / 16	-1.53 / 63	-1.57 / 61

LO 42	0.19 / 33	0.00 / 39	-0.64 / 47	-0.76 / 49	0.65 / 20	1.08 / 14
LO 05	0.17 / 34	0.26 / 33	0.91 / 19	0.86 / 19	0.77 / 19	0.69 / 21
LO 36	0.16 / 35	0.27 / 32	-0.01 / 38	0.05 / 38	-1.11 / 55	-1.35 / 56
LO 66	0.16 / 35	0.01 / 38	0.83 / 25	0.98 / 18	-0.58 / 49	-0.10 / 43
LO 54	0.15 / 37	-0.01 / 40	-0.55 / 45	-0.66 / 47	0.42 / 27	0.70 / 20
LO 08	0.08 / 38	0.03 / 37	-1.46 / 59	-1.55 / 59	0.17 / 34	0.10 / 38
LO 55	0.02 / 39	0.09 / 35	-1.01 / 51	-1.10 / 51	-1.90 / 67	-1.98 / 67
LO 57	0.00 / 40	-0.05 / 41	0.07 / 36	0.01 / 39	0.64 / 21	0.40 / 28
LO 65	-0.01 / 41	0.09 / 35	0.80 / 28	0.74 / 27	1.83 / 02	1.77 / 04
LO 29	-0.01 / 41	-0.17 / 44	0.05 / 37	0.20 / 36	0.35 / 34	0.66 / 23
LO 52	-0.17 / 43	-0.07 / 42	0.87 / 20	0.81 / 22	1.24 / 11	1.48 / 08
LO 21	-0.17 / 43	-0.10 / 43	0.92 / 18	0.86 / 19	0.63 / 22	0.69 / 21
LO 45	-0.22 / 45	-0.38 / 46	1.38 / 09	1.54 / 10	0.12 / 36	0.42 / 27
LO 35	-0.43 / 46	-0.60 / 50	0.10 / 35	0.25 / 35	0.39 / 29	0.92 / 15
LO 07	-0.45 / 47	-0.35 / 45	-0.01 / 38	-0.07 / 41	-1.43 / 61	-1.57 / 61
LO 38	-0.50 / 48	-0.39 / 47	0.27 / 34	0.46 / 32	0.01 / 41	0.24 / 35
LO 50	-0.57 / 49	-0.72 / 53	-1.11 / 52	-1.21 / 54	-0.04 / 43	0.32 / 29
LO 40	-0.60 / 50	-0.52 / 48	2.18 / 04	2.12 / 05	0.91 / 14	0.76 / 17
LO 56	-0.62 / 51	-0.55 / 49	-0.75 / 50	-0.82 / 50	0.21 / 33	0.28 / 33
LO 68	-0.72 / 52	-0.63 / 51	0.31 / 33	0.36 / 34	-0.53 / 47	-0.74 / 51
LO 14	-0.79 / 53	-0.69 / 52	-0.06 / 42	-0.11 / 42	-1.68 / 66	-1.80 / 66
LO 20	-0.91 / 54	-0.83 / 54	1.36 / 10	1.56 / 09	0.48 / 25	0.29 / 31
LO 39	-0.92 / 55	-0.83 / 54	-1.32 / 55	-1.41 / 56	-0.76 / 51	-0.65 / 49
LO 02	-0.96 / 56	-1.10 / 57	-1.97 / 64	-2.08 / 65	-0.42 / 46	-0.06 / 42
LO 03	-1.16 / 57	-1.07 / 56	-0.23 / 43	-0.29 / 43	0.58 / 23	0.49 / 26
LO 18	-1.22 / 58	-1.36 / 59	-1.38 / 57	-1.48 / 58	0.48 / 25	0.85 / 16
LO 25	-1.25 / 59	-1.18 / 58	-2.72 / 69	-2.71 / 69	-1.96 / 68	-2.10 / 68
LO 12	-1.43 / 60	-1.57 / 64	-2.37 / 66	-2.48 / 67	-2.53 / 70	-2.24 / 69
LO 15	-1.48 / 61	-1.37 / 60	0.93 / 17	0.85 / 21	0.10 / 40	0.29 / 31
LO 22	-1.60 / 62	-1.50 / 61	-0.28 / 44	-0.34 / 45	-0.73 / 50	-0.74 / 51
LO 58	-1.61 / 63	-1.53 / 63	-1.34 / 56	-1.22 / 55	-1.25 / 58	-1.76 / 65
LO 06	-1.62 / 64	-1.51 / 62	-1.72 / 60	-1.69 / 60	-1.66 / 65	-1.61 / 63
LO 24	-1.70 / 65	-1.61 / 65	-1.81 / 61	-1.90 / 62	-1.04 / 54	-0.69 / 50
LO 47	-1.76 / 66	-1.66 / 66	0.84 / 23	1.02 / 17	-0.01 / 42	0.31 / 30
LO 60	-2.27 / 67	-2.17 / 67	-1.13 / 53	-1.18 / 52	-1.40 / 60	-1.41 / 57
LO 44	-2.35 / 68	-2.23 / 68	-2.45 / 67	-2.31 / 66	-1.16 / 57	-1.43 / 58
LO 23	-2.88 / 69	-2.78 / 69	-2.08 / 65	-1.94 / 63	-1.53 / 63	-1.68 / 64
LO 59	-4.41 / 70	-4.30 / 70	-2.49 / 68	-2.54 / 68	-2.38 / 69	-2.39 / 70

5.2 Verification Study

A Learning Object Selection Study was conducted to compare the automatic selection according to EOS approach with human experts' choices. Since comparing and selecting from seventy candidate learning objects are too tedious for a human being, we needed to limit the number of learning objects that were provided to human judges to rate. From the ranked list generated for each of the three learning context, we randomly selected two learning objects from top ten (most suitable for that learning context), two learning objects from bottom ten (least suitable for that learning context), and two learning objects from the remaining fifty (medium suitability).

Human experts were asked to rate the six learning objects for a learning context according to which two were best and which two were worst. This was replicated over three learning contexts. The information supplied to human experts in the study was as much as required by the EOS approach. A learning context was described by metadata which contains values for all leaner characteristic variables (*Attitude*, *Major*, *Goal*, *Achievement*, *Programming experience*, and *Time available*) and variables for learning object evaluation (*Preferred format*, *Preferred depth*, *learner's listening level*, *Reading level*). The range of each variable was provided along with the actual value of the variable. Learning objects used in the study were represented by their metadata as well. Learning object features (*Format, Required listening level, Required reading level, Satisfied prerequisite portion, Depth, Required study time*), historical information (*teachers' recommendation*, *overall popularity*, *recommendation from students with similar academic achievement*, *recommendation from students with similar format* *preference*, *recommendation from students with similar learning attitude*, *recommendation from students with high academic achievement*), and *days in use* are all presented to the experts. In addition, the statistical data about Importance of Learning Object Features and Trustworthiness of Recommendations gathered from the Learning Preference Questionnaire were also provided to the experts.

The experts involved in the study were asked to examine three learning contexts and six learning objects for each learning context. That is, eighteen learning objects in total were rated by the experts. They needed to choose two most suitable learning objects and two least suitable ones for each learner. The entire survey document is attached as *Appendix D*.

We sent the request for participation in the Learning Object Selection Study to four experts, each of whom had teaching experience and knowledge about learning object metadata. Three of them completed the learning object selection task and associated survey. Among them, expert 1 is a university staff member with extensive teaching experience of computer science courses and broad knowledge about e-learning. Expert 2 is a graduate student, whose research focus is e-learning, having considerable computer science lab teaching experience. Expert 3 is a university staff member with some experience of teaching computer science and extensive e-learning knowledge. Among them, Expert 1 is most knowledgeable and skillful, and Expert 3 is the one having least teaching experience required for the individualized learning object selection.

5.3 Result Analysis

In this study, we consider both the human experts and machine algorithms as raters. What we are concerned with here is the correlation between rankings made by different raters, especially the level of agreement in ranking between human experts and machine algorithms. Inter-rater reliability analysis is often used for this purpose.

5.3.1 Comparison of the Two Machine Algorithms

As described in *Section 5.1.3*, all seventy simulated learning objects are evaluated for three simulated learners using both the Bayesian algorithm and the Naïve Bayes algorithm. As results, six lists with seventy learning objects for each learner – evaluation algorithm pair were generated. Each list was sorted in descending order according to the evaluation score of learning objects, and a rating value between 1 and 70 was assigned to each item. The integer number 1 was associated with the learning object having highest score (most suitable), while integer 70 was given to the least suitable one. Then three lists for the same algorithm were put together into learning object – learning context pairs so that a two hundred and ten item list, therefore, was generated by each algorithm. The two hundred and ten learning object – learning context pairs appear once in each list. Values associated with each pair were in the range 1 – 70.

The correlation coefficients between the two long lists are shown in *Table 5-7*. Spearman Rho is impressively high, which indicates an excellent correlation between the two algorithms. The statistic, Cohen's Kappa, is not very high because ranking shifts in the list are considered in a different way for this statistic.

Table 5-7 Comparison of Full Ranking Lists of Two Algorithms

Type	Number of Valid Cases	Value	Significance
Spearman Rho	210	0.982	0.000
Cohen's Kappa	210	0.208	0.000

The goal of the selection is to provide a short list of suitable learning objects for learners. The exact rank of each learning object is actually not what we are interested in. Our main concern is to distinguish the top ones from the rest. We divided the learning objects into three categories:

Category 1: represents most suitable. The top ten learning objects are considered as items in this category.

Category 2: corresponds to learning objects other than top ten and bottom ten.

Category 3: indicates least suitability. The bottom ten learning objects fall in this category.

After reassigning ranks of 1, 2, or 3 to the learning objects, the two lists of learning objects, generated by using Bayesian algorithm and Naïve algorithm respectively, are compared again. The results are listed in *Table 5-8*. Spearman correlation coefficient is greater than 0.9, and Cohen's Kappa coefficient is also above 0.9. The two algorithms, therefore, are considered highly consistent with each other.

Table 5-8 Comparison of 3-Degree Ranking Lists of Two Algorithms

Type	Number of Valid Cases	Value	Significance
Spearman Rho	210	0.924	0.000
Cohen's Kappa	210	0.904	0.000

5.3.2 Verifying Machine Selection against Human Expert Selection

Three experts completed our Learning Object Selection Study. Each one rated eighteen learning objects (six for each of the three learning contexts). The survey results, along with the machine's ranking, are summarized in *Table 5-9*. In the table, most suitable is denoted by number 1; least suitable is represented by 3; and 2 stands for medium.

Table 5-9 Learning Object Selection Study Results

Learning Context ID	Learning Object ID	Machine's Rank		Expert's Rank		
		Bayesian	Naive	Expert 1	Expert 2	Expert 3
LC00	LO06	3	3	3	3	3
	LO08	2	2	2	2	2
	LO31	1	1	1	1	1
	LO41	1	1	1	2	2
	LO54	2	2	2	1	1
	LO59	3	3	3	3	3
LC01	LO19	3	3	3	3	3
	LO27	1	1	1	1	1
	LO35	2	2	1	2	3
	LO43	1	1	2	1	1
	LO44	3	3	3	3	2
	LO68	2	2	2	2	2
LC03	LO00	1	1	1	1	2
	LO12	3	3	3	3	3
	LO13	2	2	2	2	3
	LO25	3	3	3	1	1
	LO32	1	1	1	2	1
	LO38	2	2	2	3	2

The rankings produced by the machine algorithm and different experts were compared on a pair-wise basis. *Table 5-10* shows the agreement between experts as well as the agreement between machine algorithm and each expert.

Table 5-10 Inter-Rater Agreements

Category	**Raters**	**Number of Valid Cases**	**Cohen's Kappa**	**Significance**
Machine vs Expert	Machine – Expert 1	18	0.833	0.000
	Machine – Expert 2	18	0.583	0.000
	Machine – Expert 3	18	0.417	0.012
Expert vs Expert	Expert 1 – Expert 2	18	0.417	0.012
	Expert 2 – Expert 3	18	0.500	0.003
	Expert 3 – Expert 1	18	0.333	0.046

Cohen's Kappa coefficient between Expert 1 and machine algorithm is above 0.8. It indicates that excellent inter-rater agreement exists. As mentioned before, Expert 1 is the most experienced one, and thus the judgement provided by Expert 1 should be more trustworthy. Both Expert 2 and Expert 3 have moderate agreement with the machine algorithm. Their κ values are all in the range of 0.4 – 0.6. In summary, agreement between machine algorithm and human experts is moderate to excellent.

After examining the comparison between different experts, we find that the consistency among them is even lower. The highest κ value between experts is 0.500, which suggests a moderate agreement; while the lowest is 0.333, which is associated with fair agreement.

We have to point out that human beings are notorious for their inconsistency. The same person could provide different result at different times. That is, a human expert could lack agreement with himself/herself. On the other hand, our machine algorithms would never have such inconsistent behaviour.

In summary, we conclude that machine selection is at least as good as human experts' selection of learning objects.

5.4 Summary

The optimized selection approach was tested with a simulation study. Both the Bayesian Weight Model and the Naïve Bayes Model were used in the simulated selection. The results produced by the two algorithms were compared, and the two algorithms highly correlated each other in the domain where the testing was conducted.

A Learning Object Selection Study was performed to validate the selection algorithms against human experts. By comparing machine selection and human experts' selection, we concluded that the agreement between machine selection and human experts' selection is higher than agreement among the human experts alone.

Chapter 6

Conclusion

6.1 Contributions

As stated in the earlier section of this thesis, the goal of this research was to design and develop a practical approach for dynamically selecting the most suitable learning objects for a particular learner in a web-based educational system. Three areas were explored:

- Extend existing learning object metadata specifications to meet the requirements of individualized learning object selection.
- Provide an approach to the selecting of a short list of suitable learning objects appropriate for the learner and the learning context.
- Implement a learning object selector to evaluate and validate the approach by comparing its results with human experts' judgment.

The goal was addressed in several aspects, and major efforts and contributions are summarized here.

6.1.1 Extension of Learning Object Metadata Specifications

The research started with examining existing learning object metadata specifications. Current standardization focuses were promoting reusability and interoperability. Those text-based tags for categorizing and annotating learning objects, however, do not carry enough information for individualized selection. Along with discussion about the requirements that learning object suitability assessment and individualized selection demand, suggestions were made for extending the existing specification.

The suitability of a learning object is a contextual feature. It can be decided only when the learning object is situated in a certain context. Also, some quality and appropriateness features may not be readily describable by an author or evaluator. Information gathered from prior usage of learning objects can be very helpful. To carry out individualized learning object selection, three categories of information are required to be captured and recorded.

Information about Learning Context describes learner characteristics, learning preferences, and available resource.

Information about Learning Objects is covered by existing specifications to a certain degree. Information required for educational purposes (e.g. pedagogical objective, prerequisites), however, has not been addressed sufficiently.

Information about Learning Object Usage History should include records about previous users, learners or instructors, and their interaction with learning objects. Statistical information, such as overall popularity, can be useful when there is a lack of detailed information.

6.1.2 A Model for Determining Weights

The suitability of a learning object is a contextual feature. It can be decided only when the learning object is situated in a certain context. The importance of different attributes of a learning object varies from learner to learner. The core of the optimized selection is dynamically identifying the importance of learning object attributes given a particular learner.

A Learning Preference Survey was conducted to discover and determine the relationships between the importance of learning object attributes and learner characteristics. The relation was represented in two structures, a two-layer Bayesian network and a set of Naïve Bayes Net. Either weight model can provide weights for different learning object features when given a particular learner. Those weights were used in evaluating learning objects. The resulting ranked lists of suitable learning objects produced by the two modelling approaches were highly correlated with each other. This finding could simplify the optimized selection of EOS approach.

One difficulty attached to Bayesian network deployment is determining the connections between nodes. Usually an expert is necessary. In addition, the computational complexity of inference in a Bayesian Network depends on the network structure. Inference in the underlying Bayesian Network of a student model generally tends to be very expansive. Even with various inference techniques, it still has exponential complexity in the worst case. The computational complexity of Bayesian Networks discourages their usage as well. With the Naïve Bayes approach, building a model is much simpler because the root node variable can be considered as a function of all children node variables. The computational expense becomes much less because each child node always has only one parent. Based on our results, this simpler Naïve Bayes approach seems to prove sufficient for individual selection of learning objects.

6.1.3 Individualized Selection and Validation

The Eliminating and Optimized Selecting (EOS) approach was proposed to perform individualized selection. It divided the selection procedure into two steps: eliminate irrelevant learning objects according to requirement constraints and optimize selections based on learners' characteristics and history information.

The optimized selection was explored and demonstrated using a simulation approach. Seventy simulated learning objects were evaluated for three simulated learners / learning contexts. The selection results were validated with human experts' selections via a Learning Object Selection study. By comparing machine selection and human experts' choices on pair-wise basis, we found out that the agreement between machine algorithms and a human expert is actually better than the agreement among human experts. This discovery is quite encouraging.

Another advantage that the machine algorithm has is consistency. It will always provide the same result for the same evaluation no matter how many times the task is performed.

6.1.4 Advantages of the Approach

A common challenge that a recommendation system faces is the cold-start issue. That is, when the historical usage information is not available, the selection technique does not work well. Our EOS approach for the individualized learning object selection addresses this problem by not relying on reference of "similar" users (learners). Instead, we try to build connections between learner characteristics and the evaluation of learning objects using weight models constructed in a Bayesian Belief network. Because of its capability of dealing with uncertainty, difficult situations, such as cold-start, incomplete learner data information, etc. can be handled gracefully.

6.2 Limitations and Future Work

In order to limit the scope of this research to fit within an MSc thesis project, some compromises were made that lead to limitation in the generality of the results. The work could be improved and extended in several areas.

6.2.1 Extended Domain

Since our Learning Preference Survey was conducted among first year Computer Science students studying elementary programming, the research is limited to that learning context. Also students' background was fairly uniform. The differences on many aspects that we examined were quite small. Some of them, such as registration status and first language, are so close that we were prompted to ignore them in the research. A different direction to carry this research is to try different domains, for example, Business or Medicine. With different types of learning objectives, learners' opinions on the same set of questions could very likely be different. Trying the study with learners having more variety in background (i.e. including high school students, elders) would be worthwhile as well.

6.2.2 Methodology

We employed both a Bayesian network and a Naïve Bayes algorithm to construct weight models, but variables considered were all the same. The variables that were eliminated for Bayesian Weight Model were not included in Naïve Bayes models either. Without having gone through the step of developing the Bayesian model first, our Naïve Bayes model would have been structured differently. Building Naïve Bayes model with all variables could be an alternative approach and might have resulted in a fairer comparison between the two methods. By using all variables the Naïve Bayes algorithm might have produced different outcomes.

6.2.3 Real System Exploration

The research idea was explored and tested in a simulated environment because we could not locate a sufficiently large number of similar learning objects needed to do realistic selection. It is not certain whether the conclusions we draw from simulation could well apply in the real world. The study needs to be re-run in more domains and under more realistic circumstances. Time efficiency and other practical issues could be investigated only in a real system.

6.2.4 Model Tuning

Open learner models allow for the model to be tuned by learner themselves. This would help maintain an accurate representation of learner in the system. Applying open learner model to learning object selection could be a promising new direction for future research.

6.3 Conclusion

This research investigated the problem of learning object selection. A model was developed and instantiated using data collected from real learners. The model was verified and validated using simulated learning objects and learners and human expert judges. The study showed that it may be feasible to select an appropriate learning object for an individual learner if sufficient metadata is available about the learning objects, learning context, and learner models. While the actual results of this research are not widely generalizable, the methodology is very general. In order to do learning object selection in any particular domain, an analysis of typical learners' preferences ought to be done to identify important learning object-, learner-, and learning-attributes and interconnections. Experts or prior data could generate prior and conditional probabilities to construct a weight model. Once this is done, an EOS approach is ready to support individual learning object selection in that new domain.

References

[1]. ADL Technical Team. Sharable Content Object Reference Model: Overview. 3rd Edition. Retrieved March 20, 2009, from the World Wide Web: http://www.adlnet.gov/downloads/DownloadPage.aspx?ID=237

[2]. Adomavicius, G. and Tuzhilin, A. Toward the next generation of recommender system: a survey of the state-of-the-art and possible extensions. IEEE Transactions on Knowledge and Data Engineering, 17(6). pp. 734-749, June 2005.

[3]. Brooks, C. and McCalla, G. (2006) Towards flexible learning object metadata. Int. J. Continuing Engineering Education and Lifelong Learning, Vol. 16, Nos. 1,2, pp. 50-63.

[4]. Brusilovsky, P. (2000) Adaptive hypermedia: From intelligent tutoring systems to Web-based education (Invited talk). G. Gauthier, C. Frasson and K. VanLehn (Ed) Intelligent Tutoring Systems. Lecture Notes in Computer Science, Vol. 1839, (Proceedings of 5th International Conference on Intelligent Tutoring Systems, ITS 2000, Montreal, Canada, 2000) Berlin: Springer-Verlag, pp. 1-7.

[5]. Brusilovsky, P. (2001) Adaptive hypermedia. User Modeling and User Adapted Interaction, Ten Year Anniversary Issue (Alfred Kobsa, ed.) 11 (1/2), pp. 87-110.

[6]. Brusilovsky, P. and Vassileva J. (2003). Course sequencing techniques for large-scale web-based education. International Journal of Continuing Engineering Education and Life-long Learning, vol. 13 (1/2). pp. 75-94.

[7]. Bull, S. and McCalla G. Modelling Cognitive Style in a Peer Help Network. Instructional Science 30(6), 2002. pp. 497-528.

[8]. Charniak, Eugene. Bayesian Networks without Tears. AI Magazine, Vol. 12, No. 4 (Winter 1991), 50-63.

[9]. Chen, S. Y. and Macredie R. D. Cognitive Styles and Hypermedia Navigation: Development of a Learning Model. Journal of the American Society for Information Science and Technology, 53(1): 3-15, 2001

[10]. Downes, S. Smart Learning Objects. The Learning Place. Retrieved March 20, 2009, from the World Wide Web:

http://education.qld.gov.au/learningplace/onlinelearning/courses/sdownesapril.html

[11]. Felder, R. M. Matters of Style. American Society for Engineering Education Prism, 6(4), 18-23, December 1996.

[12]. Friesen, N. Three Objections to Learning Objects. McGreal, R. (Ed). 2004. Online Education Using Learning Objects. Retrieved March 20, 2009, from the World Wide Web: http://learningspaces.org/n/papers/objections.html

[13]. Friesen, N. What are Educational Objects? Interactive Learning Environments, Vol. 9, No. 3, Dec. 2001.

[14]. Friesen, N., Fisher, S., and Roberts, A. CanCore Guidelines for the Implementation of Learning Object Metadata. Version 2.0. Retrieved March 21, 2009, from the World Wide Web: http://www.cancore.ca/guidelines/CanCore_Guidelines_2.0.zip

[15]. Greer, J. E. and McCalla, G. I. (Ed) Student Modelling: The Key to Individualized Knowledge-Based Instruction. NATO ASI Series F: Computer and Systems Sciences, Vol. 125. Berlin Heidelberg: Springer-Verlag, 1994.

[16]. IEEE 1484.12.1-2002, Draft Standard for Learning Object Metadata. The Institute of Electrical and Electronics Engineers, Inc. Retrieved March 23, 2009, from the World Wide Web: http://ltsc.ieee.org/wg12/files/LOM_1484_12_1_v1_Final_Draft.pdf

[17]. IMS Global Learning Consortium, Inc. IMS Content Packaging Specification. Retrieved March 22, 2009, from the World Wide Web: http://www.imsproject.org/content/packaging/

[18]. IMS Global Learning Consortium, Inc. IMS Learning Resource Meta-data Specification. Retrieved March 22, 2009, from the World Wide Web: http://www.imsproject.org/metadata/

[19]. IMS Global Learning Consortium, Inc. IMS Mata-data Best Practice Guide for IEEE 1484.12.1 – 2002 Standard for Learning Object Metadata. Retrieved March 22, 2009, from the World Wide Web: http://www.imsglobal.org/metadata/mdv1p3/imsmd_bestv1p3.html

[20]. IMS Global Learning Consortium, Inc. IMS Simple Sequencing Specification. Retrieved March 22, 2009, from the World Wide Web: http://www.imsproject.org/simplesequencing

[21]. Jameson, A. Numerical Uncertainty Management in user and Student Modelling: An Overview of Systems and Issues. User Modeling and User-Adapted Interaction 5 (3-4), 193-251.

[22]. L'Allier J. J. NETg's Precision Skilling: The Linking of Occupational Skills Descriptors to Training Interventions. Retrieved April 15, 2004, from the World Wide Web: http://ideal.cecs.missouri.edu/~yliu/NETg%20tech/NETginnovationcenter/prcisionskill.htm

[23]. Lemire, D., Boley, H., McGrath, S., and Ball, M. Collaborative Filtering and Inference Rules for Context-Aware Learning Object Recommendation. Technolodgy and Smart Education, 2(3). 2005. pp. 179-188.

[24]. Manouselis, N. and Costopoulou, C. Experimental Analysis of Design Choices in Multi-Attribute Utility Collaborative Filtering. International Journal of Pattern Recognition and Artificial Intelligence, 21(2), 2007. pp. 311-331

[25]. Manouselis, N., Vuorikari, R., and Van Assche, F. Simulated Analysis of Collaborative Filtering for Learning Object Recommendation. SIRTEL Workshop, EC-TEL 2007. Retrieved March 30, 2009, from the World Wide Web: http://infolab-dev.aua.gr/sirtel2007/papers/Manouselis_et_al.pdf

[26]. McCalla, G. (2004) The Ecological Approach to the Design of E-Learning Environments: Purpose-based Capture and Use of Information about Learners. Journal of Interactive Media in Education, 2004 (7) Special Issue on the Educational Semantic Web. Volume 1, p. 18. Retrieved March 23, 2009, from the World Wide Web: http://www-jime.open.ac.uk/2004/7/mccalla-2004-7.pdf

[27]. Merrill, M. D. Knowledge Objects. CBT Solutions, volume. 1, no. 11, March/April. Retrieved March 23, 2009, from the World Wide Web: http://cito.byuh.edu/merrill/text/papers/KnowledgeObjects.PDF

[28]. Merrill, M. D. Knowledge Objects and Mental Models. Wiley, D. A. (Ed), The Instructional Use of Learning Objects. 2002, pp. 261-280.

[29]. Mohan, P. and Greer, J. E-Learning Specification in the Context of Instructional Planning. In Proceedings of Artificial Intelligence in Education. 2003, pp307-314.

[30]. Mohan, P., Greer, J., and McGalla G. Instructional Planning with Learning Objects. Workshop on Knowledge Representation and Automated Reasoning for E-learning System. Acapulco, Mexico.

[31]. Ochoa, X. and Duval, E. Use of Contextualized Attention Metadata for Ranking and Recommending Learning Objects. Proceedings of 1st International Workshop on Contextualized Attention Metadata: Collecting, Managing and Exploiting of Rich Usage Information. 2006. pp. 9-16.

[32]. Pearl, J. Probabilistic Reasoning in Intelligent System: Networks of Plausible Inference. San

Mateo, California, Morgan Kaufmann, 1988.

[33]. Recker, M., Walker, A., and Lawless, K. What do you recommend? Implementation and analyses of collaborative information filtering of web resources for education. Instructional Science, 31(4-5), pp. 299-316.

[34]. Recker, M. and Wiley, D. (2001) A non-authoritative educational metadata ontology for filtering and recommending learning objects. Journal of Interactive Learning Environments. Special issue on metadata, pp. 1-17. Retrieved March 23, 2009, from the World Wide Web: http://inst.usu.edu/~mimi/papers/ile-lo.doc

[35]. Reilly, R., Kort, B. and Picard, R.W. External Representation of Learning Process and Domain Knowledge: Affective State as a Determinate of its Structure and Function. Proceedings the IEEE Artificial Intelligence in Education Conference, San Antonio, Texas, May 2001.

[36]. Riedl, J. and Konstan, J. Word of Mouse: The Marketing Power of Collaborative Filtering. New York: Warner Books, Inc., 2002.

[37]. Robson, R. Object-oriented Instructional Design and Applications to the Web. World Conference on Educational Multimedia, Hypermedia and Telecommunications, Seattle, Washington, USA, June 19-24, 698-702. Retrieved March 23, 2009, from the World Wide Web: http://citeseer.ist.psu.edu/cache/papers/cs/16139/http:zSzzSzrobby.orst.eduzSzpaperszSz2206.pdf/robson99objectoriented.pdf

[38]. Russell, Stuart and Norvig, Peter. Artificial Intelligence: A Modern Approach. Upper Saddle River, New Jersey, Prentice Hall, 2003.

[39]. Sosteric, M. and Hesemeier, S. When is a Learning Object not an Object: A First Step towards a theory of learning objects. International Review of Research in Open and Distance Learning, vol. 3, no. 2, Oct. 2002. Retrieved March 23, 2009, from the World Wide Web: http://www.irrodl.org/index.php/irrodl/article/view/106/557

[40]. Tang, T. Y. and McCalla, G. I. Smart Recommendation for an Evolving E-Learning System. In AIED 2003 Workshop on Technologies for Electronic Documents for Supporting Learning.

[41]. Tang, T. Y. and McCalla, G. Utilizing Artificial Learner on the Cold-Start Pedagogical-Value based Paper Recommendation. In Proceedings of AH 2004: International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems.

[42]. Tsai, K. H., Chiu, T. K., Lee, M. C., and Wang, T. I. A learning Object Recommendation Model

based on the Preference and Ontological Approaches. Proceeding of the Sixth International Conference on Advanced Learning Technologies (ICALT'06) 2006.

[43]. Vassileva, J. and Wasson, B. Instructional Planning Approaches: from Tutoring towards Free Learning. In Proceedings of Euro-AIED'96, Lisbon, 1996, pp1-8.

[44]. Wang, T. I., Tsai, K. H., Lee, M. C., and Chiu, T. K. Personalized Learning Objects Recommendation based on the Semantic Aware Discovery and the Learner Preference Pattern. Educational Technology and Society, 10 (3), 2007. pp. 84-105.

[45]. Wiley, D.A. Connecting learning objects to instructional design theory: A definition a metaphor, and a taxonomy. D. A. Wiley (Ed), The Instructional Use of Learning Objects. Bloomington, IN: Association for Educational Communications and Technology, USA, 2001. Retrieved March 23, 2009, from the World Wide Web: http://www.reusability.org/read/

[46]. Wolpers, M., Najjar, J., and Duval, E. (2007). Tracking Actual Usage: the Attention Metadata Approach. Educational Technology & Society, 10(3), pp106-121.

Appendix A   A Learning Preference Survey

The questionnaire for the Learning Preference Survey was drafted based on our literature survey. Domain experts were consulted, and revisions were made to the questionnaire accordingly. The Learning Preference Survey was not piloted nor validated because we focused on methods for deriving weight model from the data collected from the survey. The validity of the survey instrument was considered not to be of prime importance since the survey was used primarily to prototype a methodology for obtaining a Bayesian belief network for data.

The Learning Preference Survey was conducted online. A link to the following questionnaire was made available to students who were taking CMPT100 and CMPT111 in term 2, 2004 – 2005 regular session until at least 100 students submitted the result. Once a student logged in to the survey, the consent form approved by the University Advisory Committee was presented to the student (please find the consent form at the end of this survey document). The student was informed that submitting the survey would indicate the understanding of the study and the agreement to be a subject for the study. Students needed to use their NSID to log in, but their answers were collected anonymously. Their personal identification information was not used in the research.

<u>Learning Preference Questionnaire</u>

The purpose of this questionnaire is to find out important factors that influence your selection of learning materials. Your answers will be collected and used as anonymous records in this research. Personally identifiable information will not be used or revealed. We expect that it will take you about 20 minutes to complete this survey. An honourarium of $5 will be available to the first 100 students from CMPT 100 or CMPT 111 who complete the survey.

Note: This study is for students in introductory computer science courses, that is, CMPT 100 or CMPT 111. If you are taking both courses, please answer the questions from your perspective as a 111 student.

Part 1. Learner Background

1. What is your gender?
   - ☐ Male
   - ☐ Female

2. Are you a fulltime student?
   - ☐ Yes
   - ☐ No

3. Which year of university study are you in?
   - ☐ 1^{st} year
   - ☐ 2^{nd} year
   - ☐ 3^{rd} year
   - ☐ 4^{th} year
   - ☐ 5^{th} year
   - ☐ Above 5^{th} year

4. What is your major?
   - ☐ Computer Science

- ☐ Other Science
- ☐ Commerce
- ☐ Engineering
- ☐ Other

5. What language do you speak and read most fluently?

6. How would you rate your reading level in English?
   - ☐ Excellent
   - ☐ Very good
   - ☐ Good
   - ☐ Fair
   - ☐ Poor

7. How would you rate your ability to communicate orally in English?
   - ☐ Excellent
   - ☐ Very good
   - ☐ Good
   - ☐ Fair
   - ☐ Poor

8. What is your experience with computer programming?
   - ☐ Extensive
   - ☐ Medium
   - ☐ Limited
   - ☐ None

9. How much time are you willing to spend on your CMPT course?
    - ☐ More than 30 hours per week
    - ☐ 20 – 30 hours per week
    - ☐ 10 – 20 hours per week
    - ☐ 5 – 10 hours per week
    - ☐ Less than 5 hours per week

10. What is your academic achievement goal for your CMPT course?
    - ☐ Exceptional mark (>90)
    - ☐ Excellent mark (80 – 89)
    - ☐ Good mark (70 – 79)
    - ☐ Satisfactory (60 – 69)
    - ☐ Pass (50 – 59)
    - ☐ No particular goal

11. What's your attitude towards your achievement goal in your CMPT course?
    - ☐ Work hard to do my best
    - ☐ Do what I can
    - ☐ Don't care much

12. What is your general academic achievement in university classes over all?
    - ☐ Exceptional average (>90)
    - ☐ Excellent average (80 – 89)
    - ☐ Good average (70 – 79)
    - ☐ Satisfactory average (60 – 69)
    - ☐ Not so good (below 60)

13. Normally, what was the relation between your academic achievement and your goal in classes you took before?
    - ☐ My marks were generally higher than I thought they would be

- My marks were about the same as I thought they would be
- My marks were lower than I thought they would be
- Didn't care about my marks

14. What type of network access do you have when browsing course material?
    - Campus network
    - High Speed access at home (Cable or DSL)
    - Dial-up modem

Part 2. General Questions about Online Learning Material (Notes or other lecture material)

15. When deciding which learning material I want to study, the format in which the learning material is presented. (e.g. whether it's text, PowerPoint or PDF slides, tables, diagrams, audio, video, simulations, exercises, questionnaires, etc. or their combination) is
    - Very important to me
    - Important
    - No effect

16. When deciding which learning material I want to study, the quality of the learning material given online is
    - Very important to me
    - Important
    - No effect

17. When deciding which learning material I want to study, the speed at which I can access online learning material is
    - Very important to me

- ☐ Important
- ☐ No effect

18. When deciding which learning material I want to study, the time needed for studying online learning material is
    - ☐ Very important
    - ☐ Important
    - ☐ No effect

19. When deciding which learning material I want to study, the reading level needed to understand the online learning material is
    - ☐ Very important
    - ☐ Important
    - ☐ No effect

20. When deciding which learning material I want to study, the language comprehension needed to understand the learning material is
    - ☐ Very important
    - ☐ Important
    - ☐ No effect

21. When deciding which learning material I want to study, the gap between prerequisite knowledge of the learning material and my knowledge level is
    - ☐ Very important
    - ☐ Important
    - ☐ No effect

22. When deciding which learning material I want to study, the depth and comprehensiveness of the learning material is
    - ☐ Very important

- Important
- No effect

23. If there are factors other than those listed above that you think important, please describe:

Part 3. Learning Scenarios

Scenario 1. Suppose the introductory computer science course you are taking (CMPT 100 or CMPT 111) was an online course with no face to face lectures or tutorials, and people who registered in the course were mostly strangers. Suppose that in this online environment, there are many different learning materials available for you. The following is a list of statements about learning material selection. Please indicate your opinion for each of them.

24. The way material is presented in a course is important in my learning.
    - Strongly agree
    - Agree
    - Neutral
    - Disagree
    - Strongly disagree

25. The quality of the learning material is important.
    - Strongly agree
    - Agree
    - Neutral
    - Disagree

- Strongly disagree

26. The time needed for studying the learning material is important.
    - Strongly agree
    - Agree
    - Neutral
    - Disagree
    - Strongly disagree

27. It is important that the learning material is not presented in a way that requires a higher reading level than mine.
    - Strongly agree
    - Agree
    - Neutral
    - Disagree
    - Strongly disagree

28. It is important that there is no gap between the pre-knowledge required by the learning material and my knowledge level.
    - Strongly agree
    - Agree
    - Neutral
    - Disagree
    - Strongly disagree

29. It is worthwhile to wait for good learning material to be displayed even it takes a few minutes.
    - Strongly agree
    - Agree
    - Neutral

- ☐ Disagree
- ☐ Strongly disagree

30. It doesn’t matter that the learning material needs a higher reading level than mine if it has high quality.
    - ☐ Strongly agree
    - ☐ Agree
    - ☐ Neutral
    - ☐ Disagree
    - ☐ Strongly disagree

31. It is important that the learning material is presented in simple easily understandable language.
    - ☐ Strongly agree
    - ☐ Agree
    - ☐ Neutral
    - ☐ Disagree
    - ☐ Strongly disagree

32. It is important that learning material describes the topic in detail.
    - ☐ Strongly agree
    - ☐ Agree
    - ☐ Neutral
    - ☐ Disagree
    - ☐ Strongly disagree

33. It is not a problem at all when there is a gap between the pre-knowledge required by the learning material and my knowledge level.
    - ☐ Strongly agree
    - ☐ Agree

- ☐ Neutral
- ☐ Disagree
- ☐ Strongly disagree

34. It doesn't matter how the learning material is presented as long as it is accurate.
    - ☐ Strongly agree
    - ☐ Agree
    - ☐ Neutral
    - ☐ Disagree
    - ☐ Strongly disagree

35. I will cancel loading learning material if it takes quite a while (e.g. 2 minutes).
    - ☐ Strongly agree
    - ☐ Agree
    - ☐ Neutral
    - ☐ Disagree
    - ☐ Strongly disagree

36. The quality of learning material does not concern me as long as it is presented in an interesting way.
    - ☐ Strongly agree
    - ☐ Agree
    - ☐ Neutral
    - ☐ Disagree
    - ☐ Strongly disagree

37. It doesn't matter how much time it takes to study the learning material if it is helpful.
    - ☐ Strongly agree
    - ☐ Agree
    - ☐ Neutral

- Disagree
- Strongly disagree

Scenario 2. In the online learning environment described in Scenario 1 suppose that there are online records of various people's preferences regarding the learning materials they had used, i.e. their evaluation of the quality and usefulness of the learning materials. In your own selection of learning materials how much would you trust or rely upon:

38. Teachers' positive recommendations that the learning resource is useful
    - Highly trust
    - Somehow trust
    - Neutral
    - Somehow distrust
    - Highly distrust

39. Recommendation of students who have similar preferences on presentation format as you do
    - Highly trust
    - Somehow trust
    - Neutral
    - Somehow distrust
    - Highly distrust

40. Recommendation of students you would find easy to talk to
    - Highly trust
    - Somehow trust
    - Neutral
    - Somehow distrust

- ☐ Highly distrust

41. Recommendation of students having high academic achievement
    - ☐ Highly trust
    - ☐ Somehow trust
    - ☐ Neutral
    - ☐ Somehow distrust
    - ☐ Highly distrust

42. Recommendation of students having similar academic achievement to yours
    - ☐ Highly trust
    - ☐ Somehow trust
    - ☐ Neutral
    - ☐ Somehow distrust
    - ☐ Highly distrust

43. Recommendation of students having similar academic achievement goals to yours
    - ☐ Highly trust
    - ☐ Somehow trust
    - ☐ Neutral
    - ☐ Somehow distrust
    - ☐ Highly distrust

44. Recommendation of students having similar attitudes to learning as you have
    - ☐ Highly trust
    - ☐ Somehow trust
    - ☐ Neutral
    - ☐ Somehow distrust
    - ☐ Highly distrust

45. Recommendation of students having similar levels of prior knowledge to yours

- Highly trust
- Somehow trust
- Neutral
- Somehow distrust
- Highly distrust

46. Recommendation of students in the same major as yours

- Highly trust
- Somehow trust
- Neutral
- Somehow distrust
- Highly distrust

47. Recommendation of students in the same year of study as yours

- Highly trust
- Somehow trust
- Neutral
- Somehow distrust
- Highly distrust

48. Overall popularity of the learning resource among all students

- Highly trust
- Somehow trust
- Neutral
- Somehow distrust
- Highly distrust

Consent Form

Approved by the University Advisory Committee on Ethics in Behavioural Sciences Research on Nov 27, 2001 (BSC# 2001-198)

1. **Title of the study.**
   Learning Preference Survey

2. **Name(s), institutional affiliation(s) and telephone number(s) of researchers.**
   Jim Greer, Professor, Computer Science Department; 966-8655
   Jian Liu, MSc Student, Computer Science Department, 966-2676

3. **Purpose and objectives of the study.**
   This is an experimental study of on-line instructional support. This study is part of the research being conducted by the ARIES Group at the University of Saskatchewan, Department of Computer Science.

   The goal of the study is to find out important factors that influence your selection of learning materials.

4. **The possible benefits to the participants** will be an improved online learning environment for future users.

5. **Data Collection Procedure**
   We expect that it will take you about 20 minutes to complete this survey. Your answers will be collected and used as anonymous records in this research. Personally identifiable information will not be used or revealed.

6. **Risks or Side Effects**

It is hard to envisage any risks or side effects of the usage of the system. However, if we become aware of any such effects during the study, we will inform immediately the participants.

7. **Each participant is free to withdraw** from the study at anytime and this withdrawal will not affect the participants' academic status. If appropriate, the researcher may choose to discontinue a participant's involvement in the study. In any case data related to students who withdraw will be deleted from the study and destroyed.

8. **The anonymity** of the collected data and the privacy of the subjects would be completely protected and the information obtained from this data would be used only in theses, journal articles or conference publications written by the researchers. In any publication only aggregate data will be reported. Thus, the names and identities of the subjects would not be published in any form.

9. **The participants will be advised** of any new information that will have a bearing on the participants' decision to continue in the study.

10. If you want to acquire information on the results of the research once the study is completed, send a request to Jian Liu (jil089@mail.usask.ca)

11. Should you have any questions with regard to the study or to your rights as a participant in the research study, call Professor Jim Greer, 966-8655.

By submit the survey, it is indicated that you understand the study and agree to be a subject for the study.

Appendix B Simulated Learning Object Metadata

Table B-1 Attributes for Selection

ID	Format	Mastered * Prerequisite	Required Reading Level	Required Listening Level	Depth	Required Study Time
LO 00	simulation	0.6192	2	2	5	17
LO 01	video	0.2842	5	4	2	25
LO 02	audio	0.1079	2	2	3	26
LO 03	slide	0.5339	4	2	5	0
LO 04	video	0.7181	2	5	3	11
LO 05	simulation	0.5078	2	2	4	8
LO 06	simulation	0.0724	3	2	2	16
LO 07	text	0.3613	3	3	4	9
LO 08	table	0.4184	2	1	5	19
LO 09	table	0.6945	3	1	3	9
LO 10	video	0.9873	4	4	4	3
LO 11	table	0.6530	2	3	1	16
LO 12	video	0.0332	4	3	4	21
LO 13	table	0.1254	3	2	4	15
LO 14	text	0.2675	3	3	4	2
LO 15	video	0.1923	3	1	5	11
LO 16	audio	0.3571	3	5	3	13
LO 17	video	0.4493	1	3	3	21
LO 18	text	0.1050	5	2	5	30
LO 19	table	0.0225	2	1	1	20
LO 20	diagram	0.9392	2	2	5	1
LO 21	audio	0.9125	1	2	3	4
LO 22	exercise	0.2039	2	1	3	8
LO 23	audio	0.4927	1	2	1	6
LO 24	text	0.5341	1	1	4	11
LO 25	slide	0.8157	2	2	2	13
LO 26	table	0.4813	1	2	1	3
LO 27	slide	0.7730	3	2	2	1
LO 28	table	0.0643	2	2	1	28
LO 29	diagram	0.3050	2	2	3	27
LO 30	text	0.6559	3	1	3	6

LO 31	text	0.4612	4	2	3	15
LO 32	exercise	0.1195	1	1	3	16
LO 33	exercise	0.4105	2	1	4	24
LO 34	slide	0.8944	2	2	2	14
LO 35	diagram	0.4109	1	2	3	23
LO 36	video	0.0692	4	5	2	2
LO 37	simulation	0.9074	2	2	2	9
LO 38	diagram	0.0652	3	2	4	20
LO 39	simulation	0.4690	2	2	3	14
LO 40	text	0.7939	4	1	5	4
LO 41	table	0.2763	2	2	4	12
LO 42	audio	0.8158	1	2	4	27
LO 43	diagram	0.7744	2	3	4	12
LO 44	text	0.0406	2	2	1	2
LO 45	diagram	0.3316	3	1	4	27
LO 46	exercise	0.8561	2	2	2	2
LO 47	diagram	0.6306	1	1	5	14
LO 48	text	0.0302	4	2	4	24
LO 49	exercise	0.2961	1	2	2	8
LO 50	text	0.1198	4	1	4	24
LO 51	diagram	0.2780	2	3	2	23
LO 52	simulation	0.2198	1	2	4	1
LO 53	table	0.0389	2	3	2	10
LO 54	simulation	0.4091	2	1	3	26
LO 55	simulation	0.8900	3	3	3	19
LO 56	text	0.8814	1	2	4	1
LO 57	table	0.3220	3	2	5	9
LO 58	simulation	0.9729	2	1	1	3
LO 59	audio	0.2139	2	1	5	8
LO 60	text	0.2053	2	2	3	6
LO 61	table	0.3288	1	3	3	1
LO 62	text	0.7697	4	2	2	12
LO 63	table	0.6267	3	1	1	27
LO 64	slide	0.1363	3	1	4	26
LO 65	text	0.4276	2	1	5	10
LO 66	diagram	0.2516	1	3	3	29
LO 67	exercise	0.6922	1	2	2	6
LO 68	exercise	0.3407	3	2	2	6
LO 69	table	0.0152	2	2	1	10

Table B-2 Historical Information

ID	Days in System	Overall Popularity	Recommendation				
			Teacher	Similar Achieve-ment	Similar Format	Similar Attitude	High Achieve-ment
LO 00	478	0.6634	0.5971	0.1416	0.5570	0.8480	0.8558
LO 01	102	0.1416	0.6634	0.5570	0.8480	0.8558	0.0141
LO 02	401	0.5570	0.1416	0.8480	0.8558	0.0141	0.3538
LO 03	611	0.8480	0.5570	0.8558	0.0141	0.3538	0.2626
LO 04	616	0.8558	0.8480	0.0141	0.3538	0.2626	0.5204
LO 05	10	0.0141	0.8558	0.3538	0.2626	0.5204	0.2941
LO 06	255	0.3538	0.0141	0.2626	0.5204	0.2941	0.6192
LO 07	189	0.2626	0.3538	0.5204	0.2941	0.6192	0.2842
LO 08	375	0.5204	0.2626	0.2941	0.6192	0.2842	0.1079
LO 09	212	0.2941	0.5204	0.6192	0.2842	0.1079	0.5339
LO 10	443	0.6149	0.2898	0.2799	0.1036	0.5296	0.7138
LO 11	202	0.2799	0.6149	0.1036	0.5296	0.7138	0.5035
LO 12	75	0.1036	0.2799	0.5296	0.7138	0.5035	0.0681
LO 13	381	0.5296	0.1036	0.7138	0.5035	0.0681	0.3627
LO 14	514	0.7138	0.5296	0.5035	0.0681	0.3627	0.4198
LO 15	363	0.5035	0.7138	0.0681	0.3627	0.4198	0.6959
LO 16	49	0.0681	0.5035	0.3627	0.4198	0.6959	0.9873
LO 17	261	0.3627	0.0681	0.4198	0.6959	0.9873	0.6530
LO 18	302	0.4198	0.3627	0.6959	0.9873	0.6530	0.0332
LO 19	501	0.6959	0.4198	0.9873	0.6530	0.0332	0.1254
LO 20	711	0.9873	0.6959	0.6530	0.0332	0.1254	0.2675
LO 21	470	0.6530	0.9873	0.0332	0.1254	0.2675	0.1923
LO 22	24	0.0332	0.6530	0.1254	0.2675	0.1923	0.3571
LO 23	90	0.1254	0.0332	0.2675	0.1923	0.3571	0.4493
LO 24	193	0.2675	0.1254	0.1923	0.3571	0.4493	0.1050
LO 25	138	0.1923	0.2675	0.3571	0.4493	0.1050	0.0225
LO 26	257	0.3571	0.1923	0.4493	0.1050	0.0225	0.9392
LO 27	323	0.4493	0.3571	0.1050	0.0225	0.9392	0.9125
LO 28	76	0.1050	0.4493	0.0225	0.9392	0.9125	0.2039
LO 29	16	0.0225	0.1050	0.9392	0.9125	0.2039	0.4927
LO 30	676	0.9392	0.0225	0.9125	0.2039	0.4927	0.5341
LO 31	657	0.9125	0.9392	0.2039	0.4927	0.5341	0.8157

LO 32	147	0.2039	0.9125	0.4927	0.5341	0.8157	0.4813
LO 33	355	0.4927	0.2039	0.5341	0.8157	0.4813	0.7730
LO 34	385	0.5341	0.4927	0.8157	0.4813	0.7730	0.0643
LO 35	587	0.8157	0.5341	0.4813	0.7730	0.0643	0.3050
LO 36	347	0.4813	0.8157	0.7730	0.0643	0.3050	0.6559
LO 37	557	0.7730	0.4813	0.0643	0.3050	0.6559	0.4612
LO 38	46	0.0643	0.7730	0.3050	0.6559	0.4612	0.1195
LO 39	220	0.3050	0.0643	0.6559	0.4612	0.1195	0.4105
LO 40	472	0.6559	0.3050	0.4612	0.1195	0.4105	0.8944
LO 41	332	0.4612	0.6559	0.1195	0.4105	0.8944	0.4109
LO 42	86	0.1195	0.4612	0.4105	0.8944	0.4109	0.0692
LO 43	296	0.4105	0.1195	0.8944	0.4109	0.0692	0.9074
LO 44	644	0.8944	0.4105	0.4109	0.0692	0.9074	0.0652
LO 45	296	0.4109	0.8944	0.0692	0.9074	0.0652	0.4690
LO 46	50	0.0692	0.4109	0.9074	0.0652	0.4690	0.7939
LO 47	653	0.9074	0.0692	0.0652	0.4690	0.7939	0.2763
LO 48	47	0.0652	0.9074	0.4690	0.7939	0.2763	0.8158
LO 49	338	0.4690	0.0652	0.7939	0.2763	0.8158	0.7744
LO 50	572	0.7939	0.4690	0.2763	0.8158	0.7744	0.0406
LO 51	199	0.2763	0.7939	0.8158	0.7744	0.0406	0.3316
LO 52	587	0.8158	0.2763	0.7744	0.0406	0.3316	0.8561
LO 53	555	0.7702	0.8119	0.0360	0.3274	0.8523	0.6264
LO 54	26	0.0360	0.7702	0.3274	0.8523	0.6264	0.0257
LO 55	236	0.3274	0.0360	0.8523	0.6264	0.0257	0.2919
LO 56	614	0.8523	0.3274	0.6264	0.0257	0.2919	0.1159
LO 57	451	0.6264	0.8523	0.0257	0.2919	0.1159	0.2738
LO 58	18	0.0257	0.6264	0.2919	0.1159	0.2738	0.2152
LO 59	210	0.2919	0.0257	0.1159	0.2738	0.2152	0.0389
LO 60	83	0.1159	0.2919	0.2738	0.2152	0.0389	0.4091
LO 61	197	0.2738	0.1159	0.2152	0.0389	0.4091	0.8900
LO 62	155	0.2152	0.2738	0.0389	0.4091	0.8900	0.8814
LO 63	28	0.0389	0.2152	0.4091	0.8900	0.8814	0.3220
LO 64	295	0.4091	0.0389	0.8900	0.8814	0.3220	0.9729
LO 65	641	0.8900	0.4091	0.8814	0.3220	0.9729	0.2139
LO 66	635	0.8814	0.8900	0.3220	0.9729	0.2139	0.2053
LO 67	232	0.3220	0.8814	0.9729	0.2139	0.2053	0.3288
LO 68	700	0.9729	0.3220	0.2139	0.2053	0.3288	0.7697
LO 69	154	0.2139	0.9729	0.2053	0.3288	0.7697	0.6267

Appendix C Simulated Learner Metadata

ID	Attitude	Achievement Goal	Major	General Achievement	Programming Experience	Time Available	Reading Level	Listening Level	Preferred Format	Preferred Depth
LC 00	hard	fair	sci & eng	good	medium	limited	5	5	table	3
LC 01	not hard	excellent	commerce	excellent	none	very limited	5	5	diagram	4
LC 02	hard	exceptional	other	good	limited	medium	5	5	simulation	1
LC 03	not hard	fair	sci & eng	fair	medium	very limited	1	2	table	4
LC 04	not hard	exceptional	other	excellent	none	limited	2	4	diagram	1
LC 05	hard	fair	commerce	fair	medium	medium	5	3	diagram	4
LC 06	hard	fair	sci & eng	good	none	very limited	5	4	exercise	2
LC 07	not hard	excellent	other	excellent	none	medium	2	5	text	2
LC 08	hard	fair	sci & eng	fair	limited	very limited	4	3	video	5
LC 09	not hard	fair	other	excellent	none	very limited	3	3	video	5
LC 10	hard	excellent	other	fair	none	limited	4	5	diagram	1
LC 11	hard	excellent	commerce	fair	limited	very limited	5	3	text	2
LC 12	not hard	exceptional	other	good	limited	very limited	3	2	video	5
LC 13	not hard	excellent	other	fair	medium	limited	3	1	diagram	5
LC 14	hard	exceptional	commerce	fair	limited	limited	5	2	video	2
LC 15	hard	exceptional	commerce	good	medium	medium	3	4	diagram	4
LC 16	not hard	excellent	sci & eng	good	medium	limited	3	3	diagram	4
LC 17	hard	fair	commerce	good	limited	medium	3	5	simulation	5
LC 18	hard	fair	sci & eng	good	none	medium	3	3	exercise	2
LC 19	hard	excellent	sci & eng	excellent	none	limited	1	4	text	4

Appendix D A Learning Object Selection Study

The Learning Object Selection Study was delivered to the invited experts in person. Three of them completed the study and signed the consent form (please find the consent form at the end of this document). However the identities of the experts were not used in the research.

Learning Object Selection Study

The purpose for this study is to choose the most suitable learning object in a given learning context. There are three test scenarios in this study. All of them are about a learning concept in CMPT 100. In each scenario, a target student is described by a group of characteristics. Six candidate learning objects are listed along with their feature description and recommendation data. We assume that all learning objects are relevant to the students learning purpose. We also attach statistic data about Importance of Learning Object Features and Trustworthiness of Recommendations gathered from a Learning Preference Questionnaire conducted in 2004 among CMPT100 students.

Please examine data and choice two most suitable learning objects and two least suitable learning objects for each student.

Learning Context ID	Most Suitable Learning Objects	Least Suitable Learning Objects	Important Factors for Selection
lc00			
lc01			
lc03			

Please list factors that you think important for learning object selection but not listed in the study.

Simulated Learning Context Description

Learner ID	Attitude	Achievement Goal	Major	General Achievement	Programming Experience	Time Available	Reading Level (1-5)*	Listening Level (1-5)*	Preferred Format	Preferred Depth (1-5)**
lc00	**hard** not hard	exceptional excellent **fair**	**science & eng** commerce other	excellent **good** fair	**medium** limited none	medium **limited** very limited	5	5	**table**	3

Candidate Learning Objects

	Feature							Recommendation (%)					
Learning Object ID	Format	Satisfied Prerequisite Portion	Required Reading Level (1-5)*	Required Listening Level (1-5)*	Depth (1-5)**	Required Study Time (min)	Days in Use	Teacher	Student Overall Popularity	Student with Similar Achievement	Student with Similar Format Preference	Student with Similar Attitude	Student with High Achievement
s06	simulation	0.07	3	2	2	16	255	0.01	0.35	0.26	0.52	0.29	0.62
s08	table	0.42	2	1	5	19	375	0.26	0.52	0.29	0.61	0.28	0.11
s31	text	0.46	4	2	3	15	657	0.94	0.91	0.20	0.49	0.53	0.82
s41	table	0.28	2	2	4	12	332	0.66	0.46	0.12	0.41	0.89	0.41
s54	simulation	0.41	2	1	3	26	26	0.77	0.04	0.33	0.85	0.63	0.02
s59	audio	0.21	2	1	5	8	210	0.02	0.29	0.12	0.27	0.22	0.04

Note: * Level 1: lowest; Level 5: highest
** Level 1: least difficult; Level 5: most difficult

Simulated Learning Context Description

Learner ID	Attitude	Achievement Goal	Major	General Achievement	Programming Experience	Time Available	Reading Level (1-5)*	Listening Level (1-5)*	Preferred Format	Preferred Depth (1-5)**
lc01	hard **not hard**	exceptional **excellent** fair	science & eng **commercer** other	**excellent** good fair	medium limited **none**	medium limited **very limited**	5	5	**diagram**	4

Candidate Learning Objects

	Feature							Recommendation (%)					
Learning Object ID	**Format**	**Satisfied Prerequisite Portion**	**Required Reading Level (1-5)***	**Required Listening Level (1-5)***	**Depth (1-5)****	**Required Study Time (min)**	**Days in Use**	**Teacher**	**Student Overall Popularity**	**Student with Similar Achievement**	**Student with Similar Format Preference**	**Student with Similar Attitude**	**Student with High Achievement**
s19	table	0.02	2	1	1	20	501	0.42	0.70	0.13	0.65	0.03	0.13
s27	slide	0.77	3	2	2	1	323	0.36	0.45	0.91	0.02	0.94	0.91
s35	diagram	0.41	1	2	3	23	355	0.20	0.49	0.77	0.82	0.48	0.77
s43	diagram	0.77	2	3	4	12	296	0.12	0.41	0.91	0.41	0.07	0.91
s44	text	0.04	2	2	1	2	644	0.41	0.89	0.07	0.07	0.91	0.07
s68	exercise	0.34	3	2	2	6	700	0.32	0.97	0.77	0.21	0.33	0.77

Note: * Level 1: lowest; Level 5: highest
** Level 1: least difficult; Level 5: most difficult

Simulated Learning Context Description

Learner ID	Attitude	Achievement Goal	Major	General Achievement	Programming Experience	Time Available	Reading Level (1-5)*	Listening Level (1-5)*	Preferred Format	Preferred Depth (1-5)**
lc03	hard **not hard**	exceptional excellent **fair**	**science & eng** commerce other	excellent good **fair**	**medium** limited none	medium limited **very limited**	1	2	**table**	4

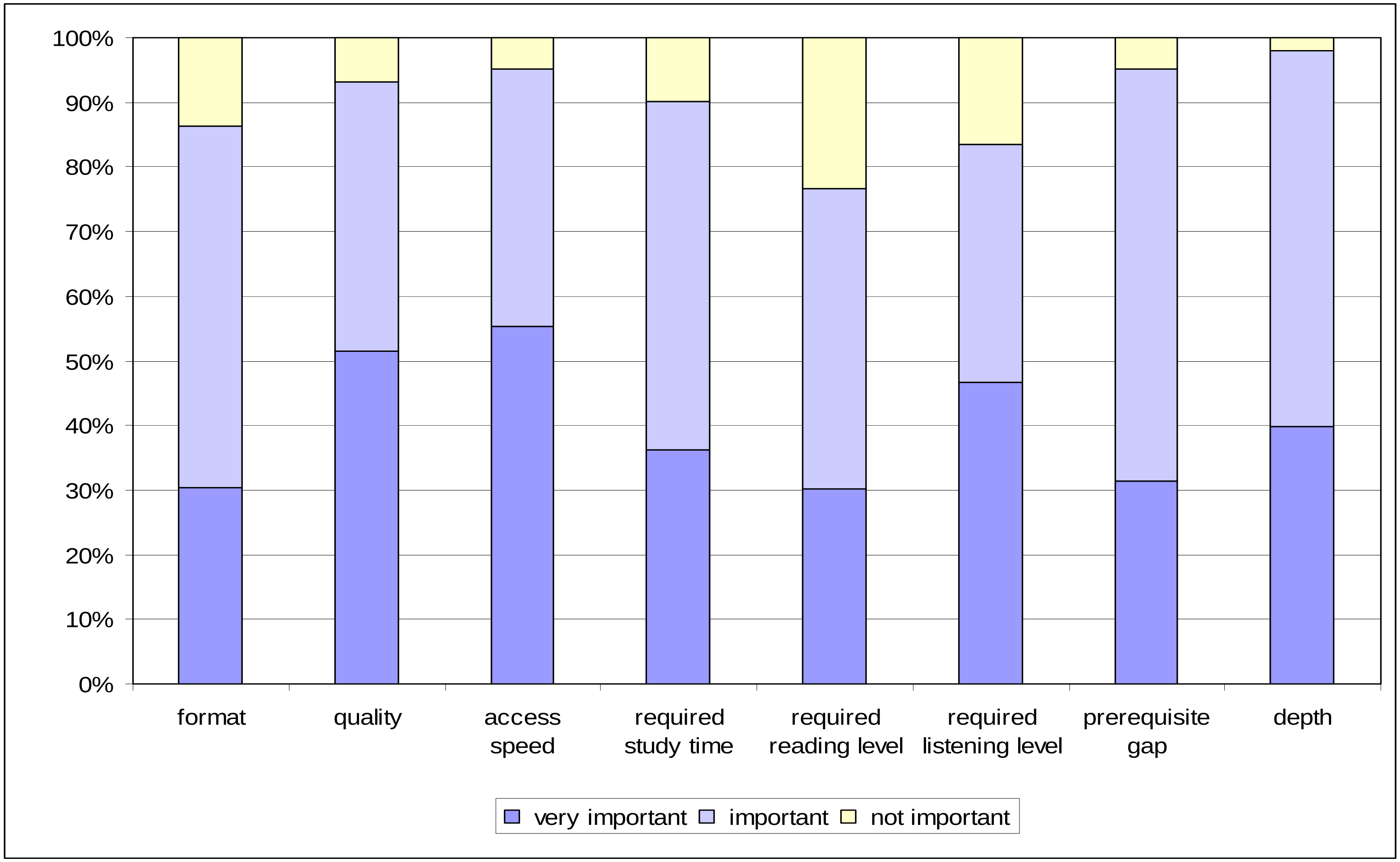
Candidate Learning Objects

	Feature							Recommendation (%)					
Learning Object ID	Format	Satisfied Prerequisite Portion	Required Reading Level (1-5)*	Required Listening Level (1-5)*	Depth (1-5)**	Required Study Time (min)	Days in Use	Teacher	Student Overall Popularity	Student with Similar Achievement	Student with Similar Format Preference	Student with Similar Attitude	Student with High Achievement
s19	table	0.02	2	1	1	20	501	0.42	0.70	0.13	0.65	0.03	0.13
s27	slide	0.77	3	2	2	1	323	0.36	0.45	0.91	0.02	0.94	0.91
s35	diagram	0.41	1	2	3	23	355	0.20	0.49	0.77	0.82	0.48	0.77
s43	diagram	0.77	2	3	4	12	296	0.12	0.41	0.91	0.41	0.07	0.91
s44	text	0.04	2	2	1	2	644	0.41	0.89	0.07	0.07	0.91	0.07
s68	exercise	0.34	3	2	2	6	700	0.32	0.97	0.77	0.21	0.33	0.77

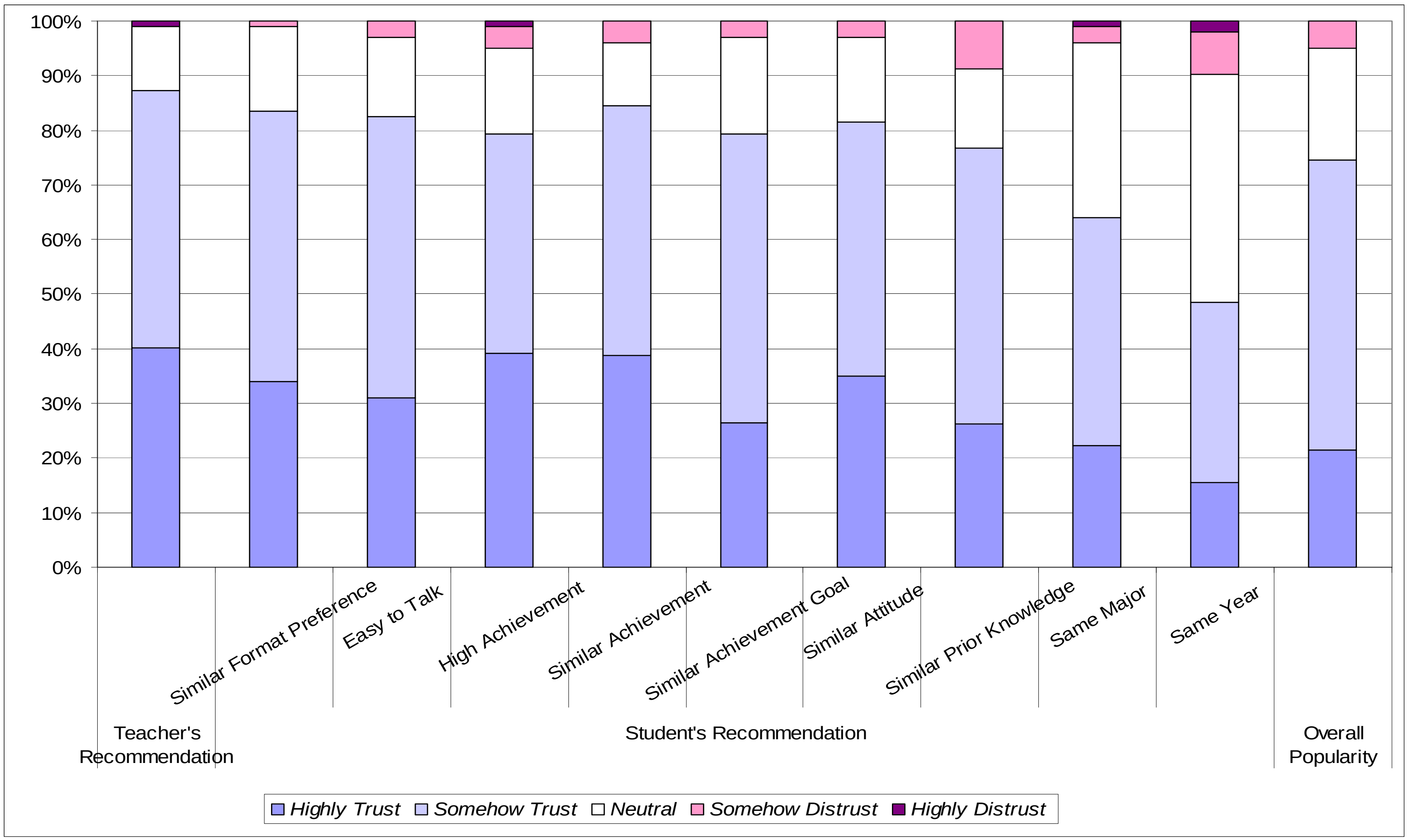
Note: * Level 1: lowest; Level 5: highest

** Level 1: least difficult; Level 5: most difficult

Importance of Learning Object Features



Trustworthiness of Recommendations



Consent Form
Approved by the University Advisory Committee on Ethics in Behavioural Sciences Research on Nov 27, 2006 (BSC# 2001-198)

1. **Title of the study.**
   I-Help: A preliminary Evaluative Study
   Learning Object Selection Study

2. **Name(s), institutional affiliation(s) and telephone number(s) of researchers.**
   Jim Greer, Professor, Computer Science Department; 966-8655
   Jian Liu, MSc Student, Computer Science Department, 966-2676

3. **Purpose and objectives of the study.**
   This is an experimental study of on-line instructional support. This study is part of the research being conducted by the ARIES Group at the University of Saskatchewan, Department of Computer Science.
   The goal of the study is to verify the EOS approach for individualized learning object selection.

4. **The possible benefits to the participants** will be an improved online learning environment for future users.

5. **Data Collection Procedure**
   We expect that it will take you about 40 minutes to complete this survey. Your answers will be collected and used as anonymous records in this research. Personally identifiable information will not be used or revealed.

6. **Risks or Side Effects**
   It is hard to envisage any risks or side effects of the usage of the system. However, if we become aware of any such effects during the study, we will inform immediately the participants.

7. **Each participant is free to withdraw** from the study at anytime and this withdrawal will not affect the participants' academic status. If appropriate, the researcher may choose to discontinue a participant's involvement in the study. In any case data related to students who withdraw will be deleted from the study and destroyed.

8. **The anonymity** of the collected data and the privacy of the subjects would be completely protected and the information obtained from this data would be used only in theses, journal articles or conference publications written by the researchers. In any publication only aggregate data will be reported. Thus, the names and identities of the subjects would not be published in any form.

9. **The participants will be advised** of any new information that will have a bearing on the participants' decision to continue in the study.

10. If you want to acquire information on the results of the research once the study is completed, send a request to Jian Liu (jil089@mail.usask.ca)

11. Should you have any questions with regard to the study or to your rights as a participant in the research study, call Professor Jim Greer, 966-8655.

The study and contents of the consent have been explained to me, I understand the contents, and that I have received a copy of the consent form for my own records.

Date:

Signatures: ______________________________ ____________________________

Participant Researcher