ABSTRACT
Self Directed E-learning focuses on the independent learner one who engages in education at his own space free from curricular obligation. The browsing history of the E-learner is used to select the document of his or her interest along with the technique of semantic query expansion.

Keywords: KNN algorithm, document retrieval, semantic query expansion, KEA key phrase extraction algorithm.

INTRODUCTION
The introduction of World Wide Web has had a profound impact on education, reducing the necessity of a learner and a teacher to share the same physical space and creating an entirely new form of knowledge delivery. With an ever-increasing number of internet users and websites, online learning, training and online educational multimedia – all generally referred to as ‘e-learning’ are becoming increasingly prevalent. E-learning systems accessible through the internet have great potential to improve education through extending educational opportunities for those who cannot use the time and place bound traditional courses and through offering new services and functions that enhance the traditional classroom. E-learning systems provide multiple ways of learning (self-placed, collaborative, and tutorial) within a common application as well as providing the potential for rich media and complex interactions.

Introduction to Self-directed e-learning
Self-directed e-learning is a facet of e-learning in which the learner is able to access a vast amount of expert-defined information, but is not necessarily subject to curricular constraints (i.e., semesters, grades etc). A common example of self-directed e-learning is of a student searching for interesting topics of research in her field. She starts out with a survey of literature pertaining to her interests and continues on to newsgroups, discussion forums, Internet search engines (like Google) and digital libraries (such as those provided by the IEEE). At the outset she may have only a vague idea of what she is actually looking for. She requires a facilitator that recognizes her interest and introduces online material that can spark further research. Tools such as Google Scholar, CiteSeer Research Index, etc. make it possible to do literature search without stepping out of one’s room. Due to the same technologies which helped make self-directed e-learning possible in the first place, these tools are in danger of delivering diminishing returns as micro-learning, lifelong education, and continuous education become the norm in our Information Age. Web Mining, however, may potentially offer a solution to this issue. Significant differences exist between students, such as their learning rate, personal interests, and prior domain knowledge. If e-learning delivery can be brought into alignment with these individual traits, the learners’ experience can be vastly improved over current models. As a specific example material could be adapted to each student pertaining to the desired knowledge or context. Designing systems that predetermine all possible usage scenarios is not feasible. It may not be practical or efficient in many situations because of the diverse and rapidly changing requirements of learners. It is necessary for an e-learning system that continually educates itself about the requirements of its learners and delivering material that is most appropriate for the individual learners. Web mining techniques have been used to help identify usage behaviour characteristics. The focus is on improving the user experience in self-directed e-learning systems through the use of Web Mining Techniques. Individualization is a key requirement and the knowledge acquiring experience can be improved by learning from previous instances. Mining data from past learning experience can provide useful insights into human learning methodology and helps in providing personal touch to the learning experience. Data mining, web mining in particular can provide immense opportunities to realize the true potential of self-directed e-learning. Web mining can also be used to aid a user by integrating the implicit information from multiple sources of web data. At the simplest level it can be a keyword oriented search. Learning is often aided by the inclusion of other kinds of data, such as a concept hierarchy and usage data. Meta information such as authors, citations and other expert defined data also help to improve the learning process. The hypothetical benefits of using web mining includes understand learner behaviour, determine the frequency of access and repeat access of learners.
to learning objects, learn who is doing what on the system, understanding dynamic behaviour of students in the web systems. Web mining can also be used to determine the e-learning system effectiveness. Web mining can provide the information about web traffic activities and patterns like which pages students always access most or least during a certain period of time. Based on the information the system administrator can adjust web cache size and web data distribution to improve the e-learning system. It will also enable university administrators and instructors to personalize online content for each student. Web mining can provide information about learning behaviours of students. Based on the formation and student’s profiles, university administrators and instructors can make recommendations for how the individual student can optimize his/her personal use of e-learning system. In the real world it is difficult to obtain reliable feedback on teaching activities. In e-learning systems students always access email, the web forum, feedback forms etc to express their concerns. Web mining can provide the quantitative feedback to instructor’s activity. Web mining tasks can be classified into three categories: web content mining, web usage mining and web structure mining.

**Web usage mining**
Web usage mining is a process of extracting useful information from server logs i.e. user’s history. Web usage mining is the process of finding out what users are looking for on the Internet. Some users might be looking at only textual data, whereas some others might be interested in multimedia data. Web usage mining is the type of Web mining activity that involves the automatic discovery of user access patterns from one or more Web servers.

**Web content mining**
Web content mining is the process to discover useful information from text, image, audio or video data in the web. Web content mining sometimes is called Web Text mining because the text content is the most widely researched area. The technologies that are normally used in web content mining are NLP (Natural Language processing) and IR (Information retrieval). Although data mining is a relatively new term, the technology is not. Companies have used powerful computers to sift through volumes of supermarket scanner data and analyze market research reports for years. However, continuous innovations in computer processing power, disk storage, and statistical software are dramatically increasing the accuracy of analysis while driving down the cost.

**Web Structure Mining**
Web structure mining is the process of using graph theory to analyze the node and connection structure of a web site. According to the type of web structural data, web structure mining can be divided into two kinds:
1. Extracting patterns from hyperlinks in the web: a hyperlink is a structural component that connects the web page to a different location.
2. Mining the document structure: analysis of the tree-like structure of page structures to describe HTML or XML tag usage.

![Figure 1: Relationship between services and users](image-url)
Literature Review
Determining the size of World Wide Web is extremely difficult. The web can be viewed as the largest data source available and presents a challenging task for effective design and access. In 2000 the size of the Web was reckoned to be some 7 million unique sites, a 50 percent increase compared to 1999. The use of web based education systems has grown exponentially in the last few years, added by the fact that neither students nor teachers are bound to a specific location. Virtual learning environments are installed more and more by universities, community colleges, schools, businesses, and even individual instructors in order to add web technology to their courses and to supplement traditional face to face courses. These systems can offer a great variety of channels and workspaces to facilitate information sharing and communication between participants in a course, to let educators distribute information to students, produce content material, prepare assignments and tests, engage in discussions, manage distance classes and enable collaborative learning with forums chats, news services etc. Self-directed e-learning enjoys the benefits of access to wealth of information via the internet, as well as knowledge of individual learning habits, so that specific needs of each individual learner can be taken into consideration. Some of the web-based self-directed e-learning applications are Google Scholar, Westlaw, LexisNexis, CiteSeer. Google Scholar is essentially a search engine for academic publications that are available online, each publication linked to each other’s by way of citations. Both natural language and Boolean searches are possible. The searches themselves can be customized to query certain authors, publications or within a year range. Google Scholar aims to rank documents the way researchers do, weighing the full text of each document, where it was published, who it was written by, as well as how often and how recently it has been cited in other scholarly literature. Westlaw is a sophisticated search and retrieval system for legal documents and other ancillary material available in Thompson West’s proprietary database. Westlaw offers a natural language search as well as Boolean search options. Westlaw is one of the primary online legal search services for lawyers and legal professionals in the United States and is a part of West, a part of Thomson Reuters. In addition, it provides proprietary database services. Information resources on Westlaw include more than 40,000 databases of case law, state and federal statutes administrative codes, newspaper and magazine articles, public records, law journals, law reviews, treatises, legal forms and other information resources. LexisNexis is a collection of information search and retrieval tools coupled with a vast library of available documents, including those related to law, academic, law enforcement, news, market intelligence, government and intelligence. LexisNexis searches offer a wide array of search options including natural language and Boolean searches, results customizable by document archive, similar result documents and selected text. Citeseer is one of the most popular online bibliographic indices related to computer science. The key contribution of the Citeseer repository is the ‘Autonomous Citation Indexing’. Citation indexing makes it possible to extract information about related articles. This eliminates the significant human effort and makes the search more effective and efficient.

Suggestion of documents to the e-learner based on K-NN algorithm.
K-NN is a type of instance based learning or lazy learning. Here the function is only approximated locally and all computation is deferred until classification. The K-NN algorithm is among the simplest of all machine learning algorithms. Both for classification and regression, a useful technique can be to assign weight to the contributions of the neighbours, so that the nearer neighbours contribute more to the average than the more distant ones.

Documents and queries are represented as vectors. Vector operations can be used to compare documents with queries. Here the training set is constructed based on the previous browsing history of the e-learner. The construction of training set can be done by KEA key phrase extraction algorithm. The training set is updated with respect to a time frame. When the user submits the new query, to the system, the query will be expanded in a semantic manner. Based on the expanded query the documents are retrieved which are closest to the training set.
CONCLUSION
This is a scalable technology which occurs in the ‘online’ phase to compute the recommendations against a possibly massive repository of educational resources in ‘real time’.

REFERENCES
[1] Ian H. Witten (1999) KEA: practical automatic keyword extraction