Chapter 1 Introduction and Statement of the Problem

1.0 Introduction

Increasingly learning is being nourished by online resources. The traditional paradigm of learning characterized by teacher-centric approaches in classrooms and textbooks are being increasingly supplemented by technology-enabled learning (TEL) [1]. The ubiquity of the Internet, affordances of digital technologies, and accessibility of digital learning resources have hastened the adoption of TEL. Learning using educational resources on the Internet offers significant advantages namely ease of access, a wide range of learning experiences is cost-effective, has best of breed practices and diversity of resources in terms of content and authorship. Additional factors for the popularity of web resources include technological support for collaboration with peers, timeliness of the information and technological support for interactivity [2][3].

Students are known to use the web to complete assignments, for knowledge discovery and self-learning. Typically, a student supplements and complements classroom learning by consumption of digital resources available on the web. Bountiful amount of learning resources with no legal barriers for personal use are located on the web. Educational content in the form of research articles, massive online learning courses (MOOC’s), tutorials, lecture notes, quizzes, videos, slides, blog articles, tools, applications, open educational resources (OERs) among other varied digital objects exists on the web [3][4]. Locating high-quality learning objects may not be all that arduous with the aid of modern best-of-breed search engines like Google. But, the “search-and-learn” approach has a couple of unpleasant repercussions. Firstly, the search results are agnostic to the specific needs and requirements of the individual learners. The search results ignore learner’s individual attributes like background knowledge of the domain, learning goals, learning styles, learner interests, learning preferences, and learning history in favor of “one-size-fits-all” approach. Secondly, as a consequence of generic search results, a learner invariably needs to expend energies to sift through resources to find the relevant ones to their requirements. Finding the relevant learning resources matching the learner requirements requires a certain amount of expertise and time on the part of a learner. Invariably this resource sieving for relevant
learning content can lead to learner frustration, disorientation in hyperspace, overburden the learner, lead the learner to run adrift in the hyperspace and finally abandon learning [5][6].

The preceding issues are mainly due to a lack of individualized results or “just-for-me” content. Thus, one can infer that more than access to content, personalized access to content is a primary barrier to learning off web resources. One possible remedy to this problem of vanilla delivery of content is personalization via adaptive hypermedia systems (AHSs). AHSs offer a curative to diverse user requirements by customizing learning in terms of “just-for-me” content delivery. Advantages of personalization include reduced learner information overload, better motivation to learn and in addition addresses the issue of learner disorientation in hyperspace. Popular examples of personalization are customized news delivery, targeted advertisements on the web, guided movie recommendations on sites like Netflix and channelized shopping on sites such as Amazon, Flipkart [6].

1.1 Motivation & Relevance

Exploring computer technology for student learning (eLearning) has been studied extensively in the developed world, with an eye on issues relevant to the developed world context. Solutions applicable to developed nations cannot be simply transplanted to the developing world due to differing cultural settings and pedagogical approaches. There is a lack of studies both in terms of quantity and quality on the use of eLearning to address the issues of the students in the developing world. Fortuitously, of late a consciousness has germinated in the developing world about the potential of freely available learning resources on the web to alter the fortunes of the lesser-haves [8]. Despite evidence of their usage in the developing world, there is an appreciable dearth of studies on the use of freely available learning resources with reference to the developing world context and problems [9].

India is projected to have the world’s largest working age population of nearly 1.1 billion by the year 2050 [10]. This implies that a huge percentage of the population is or shall be entering the schooling phase. Harnessing this demographic dividend requires well oiled social
sectors like education and healthcare\textsuperscript{1}. Regrettably, education more specifically school education is in neglect and in a steep descent. Publicly available data depicts a very sad state of affairs. Study after study has highlighted the abysmal academic standards among both the pupils and the teachers responsible for delivery\textsuperscript{2}. For example, less than one percent of eligible teachers qualified for teaching positions in the entrance exam conducted by the government of India for teacher recruitment to staff public schools [11].

Programme for International Student Assessment (PISA) is an academic evaluation conducted by the Organization for Economic Co-operation and Development (OECD) once every three years to benchmark the scholastic abilities of fifteen-year-old students across different countries. Students are tested on reading, math and science skills. One aim of the exercise is to generate data to facilitate participating countries to formulate educational policies to enhance educational outcomes in line with international standards. India last participated in PISA during the year 2011. India was ranked second to last out of seventy-four participating nations despite being represented by students of Himachal Pradesh and Tamil Nadu, states perceived to have better academic standards [12]. With the exception of a few islands of excellence like IIT’s, NIT’s and IIM’s overall picture of the educational sector is unsatisfactory. In order to address this grim scenario, there is a need to scaffold traditional delivery mechanisms and approaches to learning by exploring alternate channels to enhance educational outcomes. Figure 1.1 depicts the motivational factors for this study.

\textsuperscript{1} Human Capital Index Report, World Bank 2018
\textsuperscript{2} Annual Status of Education Report (ASER) over multiple years, accessed through http://www.asercentre.org
Acknowledging the importance of access to high-quality education for success in the knowledge economy, the National Knowledge Commission (NKC) mandated the establishment of a national repository for educational resources and their dissemination. There are strong reasons to believe that access to openly available learning resources may aid in reducing the knowledge divide, facilitate social mobility and enhance educational outcomes [7][10]. An attempt is made in this study to provide a framework for structured access to high-quality freely available learning resources to aid social mobility.

Currently, a vast majority of the eLearning systems use handcrafted content that is proprietary in nature. As a consequence, the development of systems using handcrafted content suffers from a few critical drawbacks. The construction of such systems entails huge expenses, involves large manual effort and time, requires expertise in design and execution, overloads the faculty and ignores the vast potential of readily available resources on the web to meet unique
requirements of individual users. In addition, these systems fail to **reuse** valuable learning resources and lack support for **interoperability** between eLearning systems [6], [7]. Systems developed using publicly available resources also known as open corpus resources (OCRs) aspire to use the large pool of learning resources for the construction of eLearning systems in a cost-effective manner. Although eLearning systems build using freely available resources offer a viable remedy they present their own set of challenges.

A sizeable majority of the content on the web is contributed by the developed world. The content has been developed by taking into consideration their academic and cultural needs [13]. Locating and tailoring content to the needs of developing countries like India with its numerous sub-cultures and languages entails tackling unforeseen challenges (figure 1.2) and demands expertise [14]. Despite the realization of the web as a knowledge repository being close to a decade and a half plus, the exercise of developing a system employing web resources for learning is a recent development [3][5]. Fewer than a dozen or so papers address the task of exploiting publicly available resources for eLearning. The nascent stage of developments in the domain imparts an additional hurdle due to the lack of prior work for reference (figure 1.2).

![Figure 1.2: Major challenges](image-url)
An explosion of information on the web has in its wake imparted significant challenges from knowledge discovery to knowledge utilization [15]. Although the web possesses learning objects on every imaginable topic, locating learning resources relevant to user requirements is a formidable task. By using search engines or crawlers content can be sourced easily from the web but the sourced content is agnostic to user needs. The large volume of sourced content then needs to be processed using AI techniques or algorithms for content matching user needs. As very few resources on the web are semantically annotated or possess metadata providing a description about their content, utility, or applicability processing learning resources using computers becomes difficult (figure 1.2). Every learning resource available on the web would have been created with a target audience, academic requirement following a definite pedagogical approach and presentation aesthetics. As each audience is different, repurposing of learning resources, maintaining pedagogical and aesthetic consistency poses its own challenges. The problem domain is highly interdisciplinary (figure 1.2) demanding knowledge of natural language processing (NLP), AI techniques like machine learning, big data analysis, information retrieval (IR), and educational theory among others. Finding a solution to the problem would require working with the constraints discussed above [4][5]. Figure 1.2 summarizes the major challenges discussed.

1.2 Open Learning Resources

The Internet is teeming with learning resources on every imaginable topic in sizeable quantities that are freely available with little barriers for personal use. Also, learning resources reside in digital repositories maintained by reputed organizations. These learning resources residing on the web or in publicly accessible digital repositories without financial or legal barriers for use are referred to as open learning resources. In literature content on the web are referred to as open corpus resources (OCRs) and resources hosted by digital repositories (for example Multimedia Education Resource for Learning and Online Teaching, MERLOT) are known as open educational resources (OERs). A more elaborate definition of OERs can be found in chapter two. The terms publicly available resources or open learning resources or open corpus resources have been used interchangeably in this study and they all refer to freely available resources on the web or hosted by digital repositories.
1.3 Personalized Learning

Personalization may be seen as a process of *transforming something* in order to meet the needs of a particular individual\(^3\). Personalized learning is the process of applying techniques and methods to enhance the process of individual learning by taking into consideration their attributes like background knowledge, learning goals, learning styles, and preferences [16]. On the Internet personalization is pervasive. Users encounter the process of personalization in recommendations of products or services, advertisements on sites like Amazon.com based on past interaction with the system. Even news articles are tailored to meet the preferences of the individual. Here the discussion pertains to personalization in educational settings only. One can define personalization in the educational context as a “just-for-me” service as opposed to the “one-size-fits-all” approach. The obvious advantage of personalization are the benefits offered by learner-centric methods as opposed to teacher-centric conventional approaches [17]. In this study, personalization is offered to a user by the virtue of them belonging to a class of users with similar user profiles. Personalization is undertaken at a coarse-grain level, fine-grained personalization would require a separate study in the context of publicly available resources due to lack of information about the learning objects.

Educational settings provide a range of personalization effects. Some of the ways personalization may be incorporated is by content presentation based on user knowledge levels, preference of learning media, learning paths, content sequencing, and problem-solving support among a number of others. The techniques to support personalization that may be used are collaborative, social, content-based, knowledge-based or a hybrid of these[17], [18].

1.4 Learning Process

Prior to course delivery either in traditional classroom settings or through the use of technology or a mix of both, course/learning objectives is first established. They are unambiguous statements setting out learning expectations from students after completing the course. Learning could be in terms of skills or knowledge acquisition as a byproduct of participation in the course. Teaching and learning material is then identified/developed to support and achieve the desired learning objectives. In terms of learning delivery, traditional

\(^3\) According to Cambridge Dictionary
learning settings could be used or they could be supplemented and complemented by technology. During the course delivery and post-course completion evaluation is undertaken, feedback analyzed and corrective measures are recommended from course objectives to the evaluation stage. Figure 1.3 provides an overview of the learning process discussed.

![Diagram of the learning process]

Figure 1.3: The learning process

Knowledge about the traditional approaches to learning in classrooms is common but technology-enabled learning requires some explanation. One approach to technology-enabled learning (used in this work) can be by the means of a dedicated eLearning application hosted on the web or on standalone computers. Here the content is developed by the faculty or experts for a specific course of an institution(s). Such dedicated systems for a specific purpose are known as closed corpus systems (CCSs). In order to reuse available content, eLearning applications may develop the portion of content and source the rest typically from the Internet. These systems are known as mixed corpus systems (MCSs). Then there are systems known as open corpus systems (OCSs) which source content completely from the web or other external knowledge repositories. CCSs and MCSs can meet the learning requirement completely since the complete knowledge of the content is known at system construction time. In the case of OCSs with content being external to the eLearning application, knowledge about the content is unknown at the time of system construction thereby offering numerous challenges as discussed earlier. But the
advantage of OCSs is the low cost of construction, reuse of content, availability of a wide range of content and speedy implementation.

1.5 Statement of the Problem

It is now apparent that the traditional notions of learning from teachers, and books within the confines of formal settings no longer define a student's learning activity. A couple of trends are evident; one is the increasing use of online digital resources and second is the competition from the Internet to traditional sources of knowledge. Availability of learning resources on the Internet is a welcome development but poses significant challenges in terms of resource discovery, filtering, and adaptation to meet the learning needs of the end users. School going students generally have poor searching behavior primarily due to inadequate domain knowledge. Poor domain knowledge leads to an inability to filter out unwanted information. Information seeking habits of students quickly leads to an overload of information, waste of time and consequent frustration. *Examining how publicly available educational resources also known open corpus resources can be exploited to supplement and complement traditional learning environment forms the gist of this study.* The aforementioned main problem is divided into the following objectives.

- To investigate quality rubrics pertinent to the textual format of learning resources on the Internet for identifying better quality content
- To examine the employment of readability scores as a marker of textual content comprehensibility and analyze a number of textual resources for readability score as a means to content personalization
- To explore the educational potential of videos based on their quantitative attributes and user feedback
- To propose a framework for eLearning using publicly available resources
- To construct an experimental prototype as a proof-of-concept for the evaluation of the proposed ideas

The ensuing chapters examine and attempt to address the aforementioned research objectives with a focus on the needs of *K-12* students.
1.6 Thesis Organization

Organization of the thesis and a brief snapshot of the remaining chapters follow (figure 1.4).

![Thesis Organization Diagram]

Figure 1.4: Thesis organization

Chapter two undertakes a critical examination of the literature. The chapter reviews the state-of-art techniques employed in the construction of web-based eLearning applications. Special emphasis is placed on highlighting the changing nature of sourcing learning resources in adaptive hypermedia educational systems (AEHS) and its consequent effects on the system design. Literature review then identifies four major research gaps that form the basis for incremental innovation in the domain. With reference to the identified gaps and research objectives, prior works are examined to provide a context to the study.

Chapter three examines the quality attributes pertinent to the textual format of learning resources on the Internet. Discovery of high-quality learning resources is undertaken with the assumption that such resources exhibit certain attributes reflecting the quality of the underlying
content. Textual data on the web may occur in HTML, PDF, word format etc. The textual content on the web pages contains presentation formatting code, advertisements, links and other elements that hinder drawing inferences from the text. In order to extract only the relevant textual content, a web scraper is used. A web scraper is a software that mimics a user behavior on a web page and is used to extract the textual content from web pages. On the web, the sheer volume of data available necessitates the use of a web scraper for extraction of textual content after search results are presented. The extracted textual content is further used for readability scores calculation. Continuing with the study of textual resources, the employment of readability as a marker of textual content comprehensibility is probed. One of the challenges with search results delivered by search engines is its apathy to user differences. Consequently, search results do not discriminate a query initiated by a student vs. a teacher. This lack of personalization is a major hurdle for school students in the use of the Internet. Going by the assumption that content comprehensibility is a major discriminating attribute with respect to the relevance of a web page to a user; the study examines a number of textual resources on a given topic for readability scores and attempts to predict their usefulness to the end user.

Chapter four examines the potential of selecting educational videos based on quantitative attributes and user feedback. School going students explore videos on sites like YouTube, but lack the ability to identify the suitable ones for their consumption due to poor searching skills and lack of domain knowledge to discriminate. Using techniques of sentiment analysis and quantitative attributes associated with the video, options for making video recommendations to a student are explored. In order to provide learner’s an abstraction for learning using the publicly available resources, a framework for eLearning is proposed. After expert review, the textual and video learning resources are presented to the students for eLearning via the framework. Using inputs from learning theory and inferences derived for identifying high-quality learning resources the framework is actuated.

Chapter five utilizes the experimental prototype constructed as a proof-of-concept based on the arguments presented in the preceding chapters for evaluation through two user trials and evaluation through expert analysis. Results of the experiment are analyzed, inferences drawn and reported.
Chapter six presents the outcomes of the research, limitations of the study and the scope for undertaking future work.

*Appendices* provide material that aid in understanding the user trials. They contain sample experiment evaluation forms, pre and post-trial questionnaires and a small sample of search queries used in the study.