Integrated intelligent LEARNing environment for Reading and Writing

D3.3 – System Specifications Report

Document identifier	**D3.3_System_Specs_ntua_symv_v03.docx**
Date	**2013-03-28**
WP	**WP3**
Partners	**NTUA, DOLPHIN, LBUS, UoM**
WP Lead Partner	**NTUA**
Document status	**V0.4 (FINAL)**

Deliverable Number	D3.3
Deliverable Title	System Specifications Report
Deliverable version number	V0.4 (FINAL)
Work package	WP3
Task	Task 3.3 Technical Specifications
Nature of the deliverable	Report (R)
Dissemination level	Restricted to other programme participants (**PP**)
Date of Version	2013-03-28

Author(s)	A. Symvonis,
Contributor(s)	C. Litsas
Reviewer(s)	I. Mihu

Abstract	The Report states the technical specifications for the iLearnRW system which, in conjunction with the user requirements (D3.1) and the Learning Strategies Specifications (D3.2), will be used in the design of the iLearnRW learning system.

Keywords	Hardware and Software Specifications, Operational description of iLearnRW, System Security.

Document Status Sheet

Issue	Date	Comment	Author
v01	2013-03-19	Initial draft.	A. Symvonis
v02	2013-03-21	System Specifications, Security issues	C. Litsas
v03	2013-03-25	User requirements, Client-server, Detailed System Specifications	A. Symvonis
v04	2013-03-28	Final, following Internal Review	A. Symvonis, I. Mihu (internal reviewer)

Project information

Project acronym:	**iLearnRW**
Project full title:	**Integrated Intelligent Learning Environment for Reading and Writing**
Proposal/Contract no.:	**318803**

Project Officer: Krister Olson

Address:	L-2920 Luxembourg, Luxembourg
Phone:	+35 2430 134 332
E-mail:	krister.olson@ec.europa.eu

Project Co-ordinator: Noel Duffy

Address:	Dolphin Computer Access Ltd. Technology House, Blackpole Estate West, Worcester, UK. WR3 8TJ
Phone:	+01 905 754 577
Fax:	+01 905 754 559
E-mail:	noel.duffy@yourdolphin.com

Table of Contents

1. Introduction

The aim of the iLearnRW project is to contribute towards a move away of traditional assistive software which uses a computer simply as an alternative to pen and paper and develop next generation learning software which uses a computer to facilitate the learning process for children with dyslexia and/or dysorthographia.

The key components of the iLearnRW system into achieving its aim are *user modelling*, the employment of *teaching strategies*, *content classification*, an *on-line resource bank*, *personalized content presentation*, *engaging learning activities* and *serious games*.

User Modelling: A user model for children with dyslexia will be developed, based on expert knowledge and user modelling techniques/activities. Among others, the user model will contain information regarding the errors the user is most likely to make. For each child, the iLearnRW system will build and maintain an individual profile, based on the developed dyslexia model. The interaction of the system with the user will be guided by this profile and will be customized to the user's individual needs.

Teaching Strategies: Individual users of the iLearnRW system will be supported in fulfilling their learning goals based on a teaching strategy that is adopted for each user. The teaching strategy is customized to the needs of each user, based on the user's profile, and is dynamic in the sense that it adapts to the child's progress while using the iLearnRW software.

Content Classification: The iLearnRW system supports the classification of learning material. Such classification is necessary in order to select appropriate learning material for each child, based on the child's needs, capabilities and preferences.

On-line Resource Bank: The iLearnRW system will maintain an on-line resource bank which consists of coherent collections of data supporting specific teaching strategies and being accessible to learners and educators.

Personalized content presentation: Personalized content presentation is an important feature of the iLearnRW system. The idea behind it is that, if we know the errors a child is likely to make, we can help the child read the text by enriching the text's presentation with visual cues which combine highlighting, text-reformatting and word segmentation.

Engaging learning activities: In order to be effective, the iLearnRW system is aiming toward high degree of learner engagement in any learning activity. This will be achieved through the employment of mini-games, the provision of positive reinforcement, the employment of child friendly and dyslexia sensitive graphical user interfaces.

Serious game: This is an attempt of the iLearnRW project to produce a game which, through a creative game scenario, extends the child's engagement and at the same time works towards achieving the individuals learning goals. It integrates the teaching strategy, user profile and specific learning activities.

The development of the iLearnRW system requires an understanding of the hardware and the software environment in which the system will be called to operate. This environment is specified primarily based on the desired system functionality (user requirements) but also on cost factors, the desire to use open software, the need to support data security and privacy, etc. The main objective of this document is to define the *system requirements* of iLearnRW.

2. Operational Description

In this section we present the main features on which the iLearnRW system is based. The iLearnRW's indented functionality has been described in the DoW document and is also presented in detail in the "User Requirements" deliverable (D3.1).

2.1. User Profiles

For each user the system should maintain a "user profile" which contains information about the dyslexia status of the user (types of problems, degree of severity, usual reading speed, etc.), cognitive age, interests, user preferences related to the set-up of the system, etc.

Initial profile. The initial profile a child can be either be manually built (with the help of the teacher, expert or the) based on prior knowledge about the child or can automatically built based on an "assessment". A default profile maybe also be used.

Profile Updates. Similarly to the setting of the initial profile, a profile may be updated manually by an expert or automatically by using incorporating tests and/or information related to the use of the ILearRW system by the child. In this case of a manual update, a suitable interface should be provided.

2.2. Type of Input Documents

Ideally, a learning tool for reading and writing should be able to support several document types. However, given that iLearnRW is a research project and not a commercial product, it focuses into the learning aspects of the environment and not to features that, even though they may increase its versatility, they will not create any new conceptual knowledge. For this reason, the iLearnRW system will be as simple as possible with respect to the type of documents it processes.

Plain text documents. The iLearnRW system will support loading simple text files such as .txt. Other similar formats may be considered. More complex file formats such as .pdf, even though they are extremely useful, they will not be supported since they require *optical character recognition* methods that are beyond the scope of this project. Such file formats they may be considered during the exploitation phase of the software.

2.3. Graphical User Interface

Our goal is to create a system that is highly customizable, user friendly and provides a convenient way of interaction with the user.

User friendly, simple, dyslexia and age sensitive GUI. The graphical user interface should be as simple as possible and should require minimal introduction and explanation for its use. The GUI should be appropriate for use by children (age sensitive) and specifically children with dyslexia (dyslexia sensitive). Relevant guidelines identified in D3.1 should be followed. In particular, specific care should be taken to be "text-light", utilizing pictures whenever possible.

Localized versions (English, Greek). The GUI will be available in localized versions for both languages (English and Greek) of the test-bed sites.

2.4. Text Presentation

The text should be reformatted so that its appearance should be more appropriate for children with dyslexia (based on relevant guidelines and the opinion of experts). Features of the text that are likely to change due to reformatting include: font type, font size, font colour, and line and word spacing.

Highlighting of text parts should be employed in order to help the reader focus on the part of the document being read. Several styles of highlighting have been employed in existing software, usually without significant differences between them. The system should support and extend the style of highlighting employed by the previous related project AGENT-DYSL.

The "highlighter" should be able to highlight different text parts (from a single character up to a paragraph as a unit) and to control its highlighting speed. In the event that more than two styles of highlighting are supported, the choice of style might become part of the system's "personalization" (see Section 2.5).

2.5. Personalization - Adaptive Features

Our system aims to balance the need for active participation in which children will be able to tailor aspects of their learning experience, with the requirement for a personalized learning programme. For example, by personalizing texts the system enables the children to participate more meaningfully in activities they might be typically excluded from due to their learning difficulties. Personalized text adaptation will follow a *theory-informed* design ensuring it does not challenge children's other difficulties (see D3.1).

As part of the personalization of the system to each individual user, several features can be tuned based on the child's profile (D3.1). Among others, the following features may be the subject of adaptation:

a) Language of the application (English/Greek).
b) Font type, size and colour.
c) Highlighting text at reading speed.
d) Style of highlighting.
e) Speed of highlighting.
f) Segmentation (more "severe" profiles might use more fine grained segmentation of the text). Levels of segmentation might include the paragraph-level, sentence-level, word-level (hyphenating words).
g) Adjusting the size of individual letters so that pairs of frequently confused words (dog/ god) can be easily differentiated.
h) Text analysis (Adaptation features based on text analysis).

2.6. Learning Strategies

Critical to the iLearnRW system is the incorporation of a learning strategy that will be used in helping children with dyslexia in overcoming problems with reading and writing. The learning strategy will employ several reading/writing activities, mini games, tests, selected learning resources, a serious game, etc. The strategy will be adaptive to the child's user profile and will be dynamic with respect to the child's progress. Based on the child's profile, some activities may be preferable to others, may be repeated more often, or may be executed at varying levels of difficulty. The specific dynamic individual learning program that each child follows is determined by the iLearnRW system based on expert knowledge.

2.7. Serious Games

The iLearnRW software environment will also contain a *serious game*. A serious game can be motivating, flexible and fun - thus encouraging children to engage in learning activities they might not otherwise undertake. During game play, text and learning activities encountered in other modes will appear, with learning aims integrated in game objectives. Introducing game like elements across the iLearnRW application can ensure that engagement is embedded within the entire thread of activities presented. Needless to say, the evolution of the game will be driven by the child's profile and the profile may be as well as be updated by the child's progress during the course of the serious game.

2.8. Dictionary / "Pictionary"

Dictionary. It is anticipated that the the iLearnRW system will use a dictionary. The basic usage of a dictionary component is to provide basic lexical analysis to parts of the system that are depended on words (e.g. a game that shows verbs which include the syllable "gra"). A syntactic analyser may also prove useful and used by components of the system.

A dictionary would be preferable to provide actions like spelling, segmentation, stemming, synonyms / antonyms, semantics and an "*Is it word?"* feature.

It would be also helpful if it offers the ability to execute advanced search queries. For example:

- Find words that contain a specific letter.
- Find words that contain a specific combination of letters.
- Find words that contain a specific letter in a given position.
- Find words of specific size (number of syllables or number of letters).

Pictionary. A mechanism for associating pictures with words would be also helpful and desirable. This component may be used in games and activities. The concept that it supports is that pictures can provide cues to the difficult words they are associated with.

We note that iLearnRW may just provide a framework for defining a Pictionary, that is, associating pictures to words, but not the content of such a pictionary; Building a Pictionary falls outside of the scope of the project.

2.9. Text Classification

An important functionality of the iLearnRW system will be related to text classification. As it is obvious, not all text is appropriate to be used as learning material for each user. Based on the child's age, reading skills, size of active vocabulary, type of dyslexia problems and degree of severity, some documents are more difficult than others and some may prove completely inappropriate for use. A text classification component will classify potential learning material with respect to its appropriateness for a particular child. Besides the text to be classified, input to the text-classifier will be the profile of the intended reader.

We note that a text classification component may find several uses in a learning environment. We briefly mention some of them.

Suggestion of appropriate learning material. A reading list on a specific topic a child may wish to study can consist of several texts. However, not all of them are suitable for the child's profile. The text classification component may provide suggestions on what to read. Its usefulness may become more obvious if combined with a search engine.

Ranking of files. A text-classification component may be used to support a file ranking functionality that will rank text files based on the suitability of their contents (related to reading) based on the profile of user of system.

2.10. Text-to-Speech

A text-to-speech component that will enable the iLearnRW system to provide multisensory input to its users may be incorporated. For example, the text-to-speech feature may read out difficult words to the child; instructions on how to play a game/activity may be given verbally; the text presentation may be combined with highlighting. We note that in some cases, as for example in the use combined use of a text-to-speech system with a highlighting component and customized presentation styles, we may face problems related to the quality of the speech production. In case that the quality of the produced speech is not satisfactory, the use of the text-to-speech component will be limited.

2.11. Recording historical data of iLearnRW's usage

The system should provide a feature that facilitates the data collection of historical data related to its usage. Such a feature is deemed necessary for the implementation of basic iLearnRW functionalities such as:

Support of learning activities. The interaction of the child with the iLearnRW system will be based on its profile and a dynamic personalized (based on the profile) learning strategy. In order to support the learning strategy the following information must be available:

- Detailed data on the activities attended by the child. For each activity, the learning resources that were use must be available in order to avoid reusing them (or to go over them again if required) in subsequent sessions. In addition, information related to the performance of the child in each activity must be also stored for later usage.
- Detailed data on the child's profile and its evolution while using the iLearnRW system.

Provision of feedback to teachers/experts and parents. Teachers/experts as well as parents must have access to data related to the use of the iLearnRW system, to the set of activities attended by a child, to the set of learning resources used in these activities, as well as to the child's performance . This information may be required in order to judge the effectiveness of the use of iLearnRW or in order to schedule alternative learning activities (not involving iLearnRW).

Scientific purposes. Researcher aiming to either evaluate the iLearnRW environment as a whole or to test the effectiveness of specific activities will find the historical data on the usage of the system extremely useful.

2.12. Run on a Tablet

There appears to be preference towards the use of a tablet as the main hardware device iLearnRW is deployed on. This is due to several factors:

- Tablets resembles eReader devices. They are light, easy to operate and can support a pleasant reading experience.
- Tablets are appropriate for game-like activities. Their touch screen provides a more natural way of interaction that can support increased level of user engagement during activities.
- Tablets are more convenient to use due to their size and portability. In addition, they are now powerful enough to support the basic iLearnRW functionality.
- Several low cost tablets are available in the market.

3. Categories of Data

In the current section we give a brief description of the main data categories that iLearnRW has to store and maintain. The data being stored by the system mainly concern user data (user profiles and users' history, user preferences, credential information). Furthermore, the system needs to maintain educational/ learning material used to support the operation of iLearnRW.

At a high level, the data maintained/stored by the system can be divided into three groups. We have *user data*, a *resources data bank*, and *expert data*. In the following subsections we present in detail every data category. At the end of this section, we also present a brief discussion data sensitivity in terms of security. We present a categorization that depends on how sensitive (in terms of security) each data type is.

3.1. User Related Data

- **User Profiles:** The basic information that the iLearnRW system needs to store in order to work properly is a *profile* for each child/user. It is quite important to have access to data that describe a single user and all information contained in his/her dyslexia related profile. By having access to individual profiles the iLearnRW system can provide specialized and personalized learning sessions to each of its users.
- **User Preferences:** iLearnRW will support some form of customization based on user preferences. Based on the preferences of each user, the look-and-feel of the system may be adjusted (e.g., the style of highlighting, the background color, the font type and size, etc.).
- **History:** When a user starts interacting with the system and using its components a log of its actions will be stored. By storing the history (details on each user session, the activities engaged, the learning material used and the user performance) the system could then make suggestions and maintain/adopt *learning strategies* according to each user's previous sessions.
- **User authentication data:** The iLearnRW final system interacts with multiple users (children, teachers/experts, parents). It is desirable to allow multiple users to gain access to the system from the same device (tablet, laptop, etc.). Allowing access to the system to multiple users from the same device requires an authentication mechanism. Such a mechanism is required in order to load the appropriate user profile, to load learning material for the specific user and to update the history of the appropriate user. In addition, access to user data must be controlled. For example, a teacher may be allowed to gain access only related to his/her students. In order to accomplish the above, a user authentication system and an access control system is required.
- **Statistical data:** Statistical data related to the usage of the system form each user may be directly stored or, alternatively, inferred by the historical data related to a user's learning session.

3.2. Resource Data Bank

3.2.1. Language Resources

- **Lexicons:** Since the iLearnRW system may provide word based activities (e.g., a game that displays 3-syllable words or an activity that asks from the child to spell a word), iLearnRW may include a dictionary to support these activities. The core of a dictionary is a database with the words of a language (e.g. English, Greek). It may be also helpful to store additional data

such as hyphenation rules, rules for stemming or information related to the type of each word (e.g., verb, noun, adjective, etc.).

- **Pictionary:** The iLearnRW may include activities and games that, for example, ask the user to write a word that corresponds to a displayed picture, or the system may display pictures in order to help the user recall a specific difficult word. If such functionality is to be implemented, we have to store a collection of images and associated relevant data.

3.2.2. Learning Material

- **Text:** iLearnRW's is a learning environment for children with dyslexia. This means that it has to utilize a variety of learning resources, mainly in the form of text. The texts may be provided and installed on a machine along with the installation of the software. Another option is to store them in a server so that every user can access them after a connection to the server. The advantage of the first option is that it does not require an internet connection, as the second does. The second choice though provides a more convenient way of updating the resources since one has only to update data at a central server.
- **Learning activities:** These activities can be games, supervised and unsupervised exercises, tests, etc. In a sense, each activity is an executable program that may require specific data in order to be carried out. In most cases, the content of each activity will be user dependent (based on the user's profile and history).

3.3. Expert Data

- **Learning Strategies:** The experts' knowledge on how to structure a strategy to facilitate learning reading and writing for a child with dyslexia must be encoded within the iLearnRW system. At the moment, it is not clear how this knowledge will be represented; this is a decision to be made during the course of the project.

3.4. Data Sensitivity

We attempt to classify the data types described above into two groups according to their sensitivity with respect to security.

- **Low Sensitivity:**
  - Language Resources
  - Learning Material
  - Expert Data
- **High Sensitivity:**
  - User Related Data

Note that language resources, learning material and expert data are user independent. It suffices to store them in a server or in every device that runs the iLearnRW software. So, no need for extra data protection is required.

On the other hand, all types of data that are related to specific users are classified as critical. Profiles must be protected since they are personal sensitive data for each user. The same holds for the history log and the preferences of each user. Also, the importance of protecting authentication date is straightforward.

4. A Client-Server Architecture

The convenience of using a server has already been indicated in the previous sections. Here we describe and briefly analyse the usefulness of implementing a client-server architecture to support various aspects of the iLearnRW system.

4.1. Server Advantages

A central server can provide to the iLearnRW the following functionality:

Ability to store a large data collection (Resource Data Bank). If a server is included in the system, then we can store a variety of text, images and/or sounds on it. The stored material can be used by the applications that run on the client side. By adopting this approach, the available databases can be larger than the database that could be stored on a single tablet. It also provides us the ability to make regular update to the databases regularly which immediately become available to all users.

Centralized control of the users' data and progress. All user related historical data and statistics can be stored (most likely in encrypted form) in the central server in order to be accessible from authorized users of the system. A GUI may be developed (client side) in order to give access to the teacher/expert/parent that is responsible for each child. They can monitor each child's progress and the activities that it engages in. The teacher/expert can make adjustments to the child's learning strategy in order to provide him/her with activities that are more relevant to his/her profile.

Single user – multiple devices. Provided that user's data have been saved on a remote server then a user could access the system from multiple devices (tablets). This is possible to be done since the iLearnRW components can be built so that on their start-up they load the profile and all required history data from the server.

Regular profile updates. By saving the user profiles and all historical data on a remote server, a teacher/expert is able to update a profile when he/she sees that a child's performance (based on his/her history) has changed. The updated profile will become available to all machines running iLearnRW or accessing its resources.

Perform expensive tasks to the server rather than the client side. A server that runs on a powerful machine can be helpful and spare the client side (i.e., a tablet) form executing difficult and resource demanding tasks. An example of this usage is the dictionary queries that will take place throughout a client session. If the dictionary is stored on the server, we can develop special software (including a web based API) that can provide to a client fast access to it.

Space, Time Trade-off. By including a powerful machine as a server in our project we have the ability to reduce the cost of some demanding tasks that are taking place throughout a client session. This can be achieved by pre-computing and storing (maybe large) data in our server that can help it perform fast computations on specific problems that arise at the client side.

User Authentication. At its simplest form, user authentication is required in order to load the correct profile for each child using the iLearnRW system through a tablet. It can be also used as an access control mechanism giving access to each user only to the data he/she is entitled to. Data necessary for user authentication are sensitive and require appropriate data protection measures. For the passwords, only their cryptographic hashes will be stored.

4.2. Web-Based Applications

A web server offers the possibility of developing software that is centrally controlled. Furthermore, web services that run at a server and take advantage of the extra processing power (or , possibly, of a larger database) may be developed. It is envisaged that the iLearnRW environment can include applications that run as a web application in order to provide to the teachers, experts, and parents rapid and easy access to children's data as well as to all available education resources of the online resource bank. These web applications can be run at the tablet's browser but also from a browser running on a laptop or desktop PC. This can be more efficient since laptop and desktop PCs have more efficient file system and provide easier access to word-processors and printer devices.

Examples of web services that the iLearnRW may use include:

A dictionary. A dictionary on consists from a large collection of words and from a set of services (complex data structures and algorithms) supporting fast and complex search queries.

Learning strategy data management. A web application that provides access to teachers, experts, and/or parents to data related to a child they are responsible for. Through this application the learning strategy for a child may be modifies and/or extra learning material may be specified and uploaded to the server.

5. System Specifications

From the previous sections we have indicated that the iLearnRW system may be considered as a client/server application. The data may be stored in a central "project server", or a server maintained by the corresponding project partner, in the form of user profiles (and, possibly, data used to construct them). These profiles will be password and/or IP address protected. Thus, no unauthorized person in the scope of the project will have access to these data. A child should be able to use any compatible device (at school or at home) that has the iLearnRW software installed (client).

Basic indicative information exchanges between client and server include:

a) User logon on server (Client → Server).
b) Server acceptance or rejection of the user connection (depends on login/password and IP address) (Server → Client).
c) Corresponding profile is loaded (Server → Client).
d) Information flow from a computer to server (updating a user profile) (Client → Server).

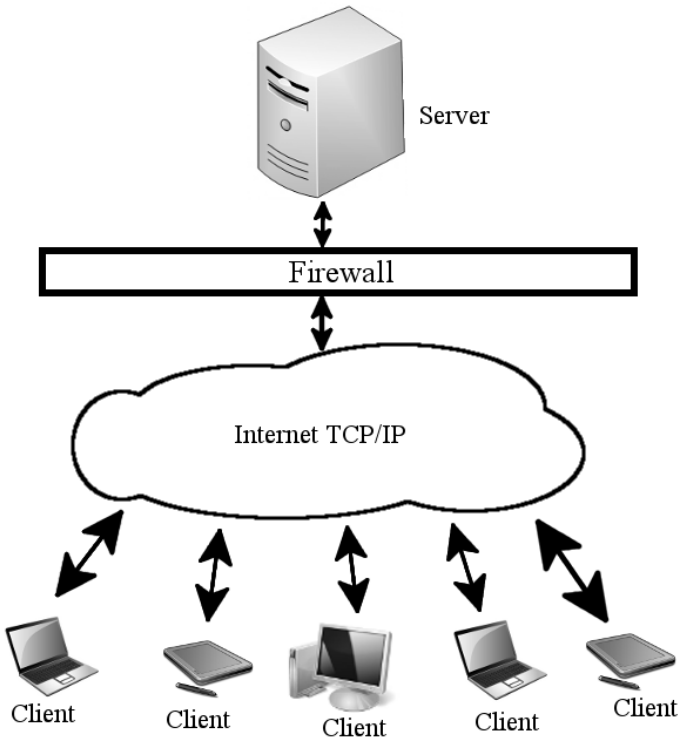

Figure 1: Server/Client Architecture

5.1. Hardware – Software Specifications

Given the high-level description of the client/server system given above, the following sections contain the server and client specifications. The description is restricted only to features that concern the iLearnRW system.

5.1.1. Server

Our goal is to set up a server which is of low cost and is also easily maintainable. We choose to use mainly free and open source software with support from large communities. In the next list we first describe the software specifications:

Software Specifications:

- **Linux based OS:** The operating system is the "core" of every server. The alternatives are to build a server based on either a MS Windows operating system or on a Linux-based operating system. A Linux-based operating system seems to best fit our needs. The main reason is that the Linux OS is supported by a wide community and it is open source and free.
- **Apache Server:** The server software is the component that allows a machine to share part of files resources over the web. At this moment, the Apache server appears to satisfy the requirements of the iLearnRW's server. The software is widely used, is open source and supported by a wide internet community. In addition, it is free to use.
- **MySQL:** It is likely that the iLearnRW system will utilize a database in order to organize and store its data on the server. The MySQL relational database seems to fulfil the system requirements. MySQL is also a free and easy to connect with the other components of the server.
- **PHP:** The iLearnRW's environment may allow to some of its users (mainly teachers/experts/ parents) to connect to the server through a web application (e.g. a teacher that wishes to check the children progress). PHP is an open-source, server-side, scripting language designed for Web development and the production of dynamic web pages. With the use of the PHP the system could have interactive web pages and easy user connection to a database or to other files/parts of the server.
- **JRE:** In many cases a client-server application includes web-services built in java in order to provide APIs for interaction between the clients and the server. The Java Runtime Environment (JRE) provides the libraries, the Java Virtual Machine, and other components to run applets and applications written in the Java programming language.

In order to provide support of multiple connections at the same time it is required that the server contains the suitable hardware. We choose a fast processor, and plenty of RAM for internal storage. A list with the basic hardware components follows:

Hardware Specifications:

- A modern fast processor (e.g. Intel Core i-5 and up series)
- Enough memory in order to run the operating system and the (e.g. 8 GB RAM)
- Disk storage sufficient to store the data (e.g. 1 TB)
- Network card.
- Mass storage device or secondary hard disk to be used for backup purposes.
- USB Ports.
- CD / DWD writer.

5.1.2. Client (Tablet)

At this time, it appear that the tablet is the device most likely the main end users (i.e., children with dyslexia) will utilize in order to gain access to the iLearnRW environment. Given that iLearnRW is a research project that has to prove a concept, it is not important to develop a system that runs on all commercially available tablet platforms.. We are going to implement the client software in Android devices. This choice was made because Android is an open source system which provides a variety of free and high quality development tools, libraries and APIs. In addition, there is available in the market a significant collection of reliable and low cost Android tablets.

It is desirable that a tablet fulfils the following requirements in order to be suitable for the project's purposes:

Tablet Specifications:

- Android OS
- A fast CPU
- At least 512 MB of RAM
- At least 1 GB of available storage (SD card)
- Touch Screen
- Screen Size 7 to 10 inches
- Screen Aspect Ratio 16:9 or 3:4
- Screen Resolution, at least 600 pixels height and at least 800 pixels width
- WiFi connection support (802.11 b/g/n)
- USB/Mini USB Ports (at least one)
- Supported audio formats: WAV, MP3, MIDI
- Embedded speakers (and option for jack)
- A battery that lasts at least 2 hours.

Tablet Desired Specifications:

- 3G internet connection
- Sensors
    1. Gyroscope
    2. G-sensor
    3. Tilt Sensor

5.1.3. Internet Connection

There are two possible ways for a user to access the web. The first is to have direct access to it. The problem with this approach is that the user may have unlimited access to every web site. This could act negatively since the user may lose his/her focus. The second option is to provide restricted internet access to the user. This could be done if the user first connects to a server that is controlled by the iLearnRW members. Then the server will re-directing the clients only to permitted pages.

In case that the final product runs on tablets there are also two options to reach the internet. A tablet can access the internet through a Wi-Fi connection or a 3G connection.

5.1.4. Backup

In the final system (client/server application) the users' profiles and related user data will be stored in a central server, so each child will be able to use the iLearnRW software from different devices.

Profiles record the children's reading progress, so it is important to keep backups. We will use the standard facilities of the Linux operating system to backup profiles and user data.

We present two simple ways for backup:

a) Backup can be a simple process of data copy in a secondary hard disk or CD/DVD. The copy process can be done "on demand" or automatically (e.g., scheduled twice a week).
b) A "full" server backup will include also the iLearnRW data. Data Backup of the central server (profiles etc.) is an obligation of the server administrator.

5.1.5. Interoperability

Interoperability is the ability of a system to provide services to and accept services from other systems and to use the exchanged services so that the systems effectively work together, e.g. a research oriented application use system data for specialized analysis. In order to achieve high degree of system interoperability, the use of de facto protocols and models of storage is required. The last is also important for treatment and transmission of information. More precisely, the following are required:

a) A pre-determined form for information storage (standards of data and meta-data formalization).
b) A pre-determined way for information exchange (communications and protocols technologies where information is transmitted according international standards (a text file with standard format, etc)).
c) A pre-determined way for data access.
d) A pre-determined way for data organization (technologies of metadata etc.).

5.1.6. System Security

The data security issues in the final system are those of a typical client/server application. The issues concern the unauthorized access of the server and the transmission of application data. Both of these issues will be addressed applying standard available software solutions (use of encryption/decryption techniques, allowing access only to specific client server applications, firewalls, etc). The exact mechanisms to be employed will be decided during implementation time and will depend on design decisions concerning the architecture of the software system. In the next section, we present in detail the basic security issues and a list of the relevant current available technologies.

6. Basic Security Issues

In the frame of the project, the iLearnRW team is called to study and to implement a completed security solution for the system. For this reason, the general security issues that are taken into consideration will be reviewed.

We define the term "Security" by the following six keywords:

- **Confidentiality:** "... is a set of rules or a promise that limits access or places restrictions on certain types of information " (http://en.wikipedia.org/wiki/Confidential)
- **Availability:** "Assuring information and communications services will be ready for use when expected." (www.ee.oulu.fi/research/ouspg/sage/glossary)
- **Integrity:** "Protection of the information from unauthorized access or revision, to ensure that the information is not compromised through corruption or falsification." (http://www.defense.gov/pubs/definitions2.html)
- **Liability:** For every transaction over the network a unique pair of sender and receiver. Neither the sender nor the receiver can dispute the transaction.
- **Authentification:** The identities of the sender and the receiver of data are assured.
- **Accessibility:** The requested data and services are always accessible.

6.1. Security Requirements of iLearnRW Systems

Beside the normal requirements to security as described in the section above, the iLearnRW system has some special requirements that must be considered during its design. In detail, a secure iLearnRW system provides:

- **Integrity of Data:** Sensitive data that are exchanged through the network are not altered.
- **Confidentiality:** Data exchanged through the network can only be accessed by authorized devices and persons.
- **Authentication:** A secure network mechanism exists for determining the identity of the source or the destination of a communication.
- **Accessibility:** All information and communications facilities are always accessible. An attacker inside or outside the network may not be able to deter normal users from accessing data within the network (denial-of-service etc.).
- **Secure File Transfer:** If an exchange of traceable personal data among the different research sites is required, secure transition mechanisms will be employed (for example, secure FTP).
- **Liability:** For every transaction over the network a unique pair of sender and receiver should be identifiable. Neither the sender nor the receiver can dispute the transaction. This is especially important in conjunction with file transfer.
- **Protection from Viruses:** The system must be secure enough in order to be protected from viruses, worms, trojans, etc.

6.2. Authorization/Operation Plan

The authorization Plan should determine "roles". For each "role" concrete rights of access should be specified. We note that authorization might be achieved for access through the client application or directly at the server. Children and researchers will use an authentication mechanism. The teacher/parent might use an authentication mechanism or they can use facilities hidden from a child (for example, the teacher/parent can print a child's statistics) despite the fact that it has not been

specified yet whether the system will use an authentication mechanism, we describe what authentications/operations the system might be able to support:

- **Child:** An authentication mechanism will be used in order a) to allow access to the system and b) to ensure that the appropriate user profile is loaded.
- **Teacher/Parent:** The teacher might be authorized (use of a login-name and a password) to help the child with the authentication (for example, by knowing the login name and password) so that the child uses the system, even when he/she has forgotten the authorization details. The teacher may wish to obtain information regarding the child's interaction with the system and to utilize this information in order to produce related activities (possibly outside iLearnRW) appropriate for the child (based on the child's profile).
- **Researcher:** Moreover, authentication (use of a login-name and a password) will be required by researchers to access stored data on different research sites. If an exchange of traceable personal data among the different research sites is required, secure transition mechanisms will be employed (for example, secure FTP, or explicit encryption/ decryption).

6.3. Current Technologies for Security

In this section, we present available technologies for security of client/server systems. The exact security mechanisms to be employed will be decided during implementation time and will depend on design decisions concerning the architecture of the software system.

SSL (Secure Sockets Layer): Secure Sockets Layer (SSL) is a protocol globally accepted for authentication and encryption of communications between clients and servers. The SSL runs above transport layer protocols like TCP/IP and below application level protocols (HTTP, IMAP) whilst providing authentication by employing public-key encryption. The SSL protocol is constituted from two sub-protocols namely: the SSL Record protocol and the SSL handshake protocol. The first is responsible in defining the format of the data to be transferred while the second uses the SSL Record Protocol for exchanging a series of messages between the client and the server in the first connection establishment. The handshake protocol is employed before any data transmission and allows the server and the client to authenticate each other. The main security issues that SSL addresses server authentication, client authentication and encrypted connectivity.

SSL Server authentication mechanism: A SSL enabled client application can employ known public key techniques to confirm whether the certificate and the ID of a server are valid and to check if the server's certificate authority is listed among the list of the client's trusted certificate authorities.

SSL Client authentication mechanism: A SSL enabled server application employs the same techniques as the ones for the server authentication, to confirm whether the certificate and the ID of a client are valid and to check if the client's certificate authority is listed among the list of the server's trusted certificate authorities.

SSL Connection (Encrypted): Confidentiality is important to any private transaction. Confidentiality in a client-server communication is achieved by the encryption of the transitive information by the sending application followed by the decryption of this information by the receiving application. In a SSL connection a complementary mechanism exists, which automatically detects if any alteration to the transferring data have been occurred.

SSH (Secure Shell): In computing, Secure Shell or SSH is a set of standards and an associated network protocol that allows establishing a secure channel between a local and a remote computer. It uses public-key cryptography to authenticate the remote computer and (optionally) to allow the remote computer to authenticate the user. SSH provides confidentiality and integrity of data exchanged between the two computers using encryption and message authentication codes (MACs). SSH is

typically used to log into a remote machine and execute commands, but it also supports tunneling, forwarding arbitrary TCP ports and X11 connections; it can transfer files using the associated SFTP or SCP protocols. An SSH server, by default, listens on the standard TCP port 22.

Secure HTTP: The Secure HTTP (S-HTTP) is an extension to the HTTP protocol designed to enable the exchange of secure data over the Internet. The difference between SSL and SHTTP is that while the first protocol is designed to provide secure connection between the sender and the receiver, SHTTP is designed to send individual messages securely. Apart from the various mechanisms offering confidentiality, integrity and authentication, SHTTP provides three other features. The first is the message encapsulation, with which an encapsulated message can include signing, based authentication and encryption. The second is that it includes headers, which provide definitions about certificate and key transfers and other communication administrative functions. The last feature is that it allows the user's involvement as well as the user's supervision to encryption and authentication activities.

FTP: FTP or File Transfer Protocol is used to connect two computers over the Internet so that the user of one computer can transfer files and perform file commands on the other computer.

The *original* FTP specification is an inherently insecure method of transferring files because there is no method specified for transferring data in an encrypted fashion. This means that under most network configurations, user names, passwords, FTP commands and transferred files can be "sniffed" or viewed by anyone on the same network using a packet sniffer. This is a problem common to many Internet protocol specifications written prior to the creation of SSL such as HTTP, SMTP and Telnet. The common solution to this problem is to use either SFTP (SSH File Transfer Protocol), or FTPS (FTP over SSL), which adds SSL or TLS encryption to FTP.

- FTP over SSH**:** FTP over SSH refers to the practice of tunneling a normal FTP session over an SSH connection. Because FTP uses multiple TCP connections (unusual for a TCP/IP protocol that is still in use), it is particularly difficult to tunnel over SSH. With many SSH clients, attempting to set up a tunnel for the control channel (the initial client-to-server connection on port 21) will only protect that channel; when data is transferred, the FTP software at either end will set up new TCP connections (data channels) which will bypass the SSH connection, and thus have no confidentiality, integrity protection, etc. If the FTP client is configured to use passive mode and to connect to a SOCKS server interface that many SSH clients can present for tunnelling, it is possible to run all the FTP channels over the SSH connection. Otherwise, it is necessary for the SSH client software to have specific knowledge of the FTP protocol, and monitor and rewrite FTP control channel messages and autonomously open new forwardings for FTP data channels. Version 3 of SSH Communications Security's software suite, and the GPL licensed FONC are two software packages that support this mode.

  FTP over SSH is sometimes referred to as secure FTP; this should not be confused with other methods of securing FTP, such as with SSL/TLS (FTPS). Other methods of transferring files using SSH that are not related to FTP include SFTP and SCP; in each of these, the entire conversation (credentials and data) is always protected by the SSH protocol.

- FTPS: FTPS (commonly referred to as FTP/SSL) is a name used to encompass a number of ways in which FTP software can perform secure file transfers. Each way involves the use of a SSL/TLS layer below the standard FTP protocol to encrypt the control and/or data channels. It should not be confused with SSH file transfer protocol. The most common uses of FTP and SSL are:
  a) AUTH TLS or Explicit FTPS, named for the command issued to indicate that TLS security should be used. This is the preferred method according to the RFC defining FTP over TLS. The client connects to the server port 21 and starts an unencrypted FTP

session as normal, but requests that TLS security be used and performs the appropriate handshake before sending any sensitive data.

b) Implicit FTPS is an older, but still widely implemented style in which the client connects to a different port (usually 990), and an SSL handshake is performed before any FTP commands are sent.

Firewall: A firewall is a mechanism used to protect a private network from another unsecured network by preventing uncontrolled access. In more detail the firewall is a device through which all traffic must be passed, keeping unauthorized users out of the private network, and providing various kinds of protection. Mainly there are three types of firewalls, the Routers, the Application-level Gateways and the Circuit-level Gateways.

- **Routers:** Routers apply rules to incoming IP packets to forward or to discard the packet. A typical router is configured to filter incoming and out coming packets (from and to the unsecured network). The rules that a router applies are subject to source and destination IP addresses as well as the TCP port number. Thus, the filters are sets of rules based on matches in the IP and TCP headers. Based on this, whenever there is a match to one of the rules then this rule is employed for making the decision whether to forward or to discard the packet. In the case where there is no matching to any rule, then default actions are taken. The two default actions are the discard and the forward, subject to router's configuration.

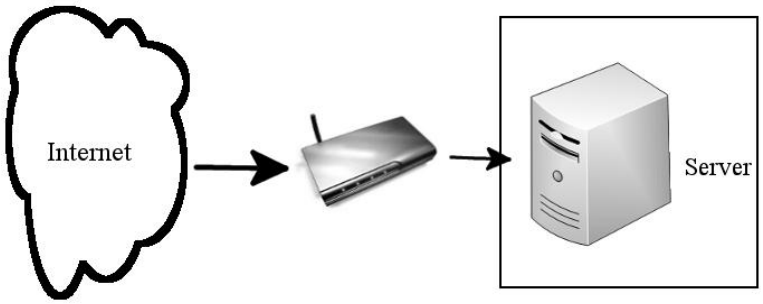

Figure 2: A Router

- **Application-level Gateway:** The application-level gateway can be thought as a switch that contacting the application-level communication between the two end-points. When the user contacts the gateway (using a TCP/IP application) is asked to give the name of the remote host to be reached. If the user provides the valid authentication information then the gateway contacts the remote host application and transfers the TCP segments containing the application data between the points. If the gateway is unable to implement the proxy-code for a specific application then the service cannot be forward through the firewall. The advantage that the gateway has over the gateway is that it needs only to examine the allowable applications while the router has to deal with number of possible combinations at the TCP and IP levels. The disadvantage of the application level gateway is that additional processing is required on each connection.

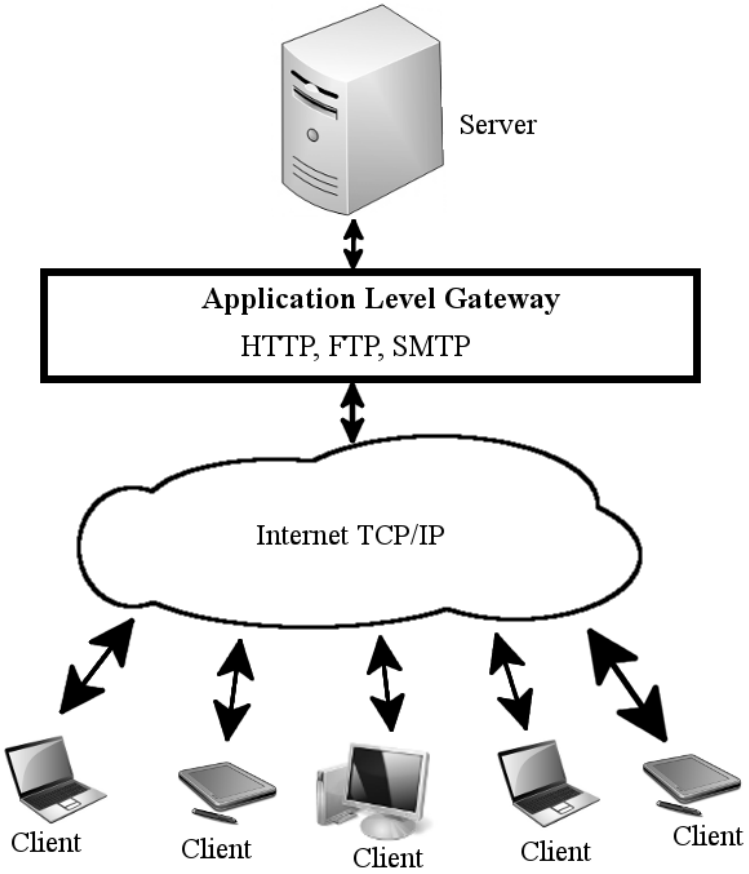

Figure 3:Application Level Gateway

- **Circuit-level Gateway:** Circuit-level gateways can be thought as special function of application level gateways. This kind of firewall does not permit end-to-end connections, instead it establishes two connections, one with the TCP user of the inside host and one with the user of the outside host. When these two connections are set up the gateway sends the TCP segments from the one connection to the other without examining the contents, while the security mechanism determines which connections will be allowed. Hence we can conclude that typical use of this kind of firewall is in cases where the system administrator trusts the internal users.
- **Proxies:** A proxy device (running either on dedicated hardware or as software on a general-purpose machine) may act as a firewall by responding to input packets (connection requests, for example) in the manner of an application, whilst blocking other packets. Proxies make tampering with an internal system from the external network more difficult and misuse of one internal system would not necessarily cause a security breach exploitable from outside the firewall (as long as the application proxy remains intact and properly configured). Conversely, intruders may hijack a publicly-reachable system and use it as a proxy for their own purposes; the proxy then masquerades as that system to other internal machines. While use of internal address spaces enhances security, crackers may still employ methods such as IP spoofing to attempt to pass packets to a target network.