An Intelligent and Contextual Information Retrieval Environment for Lifelong Learning

Luis Carlos Costa Fonseca, Crediné Silva Menezes, Rosa Vicari, Jonatas Soares
lccf@deinf.ufma.br, credine@inf.ufes.br, rosa@inf.ufrgs.br, jonatas.soares@globo.com

Abstract - The current understanding of Lifelong Learning (LLL) does not refer simply to recurrent or adult education but encompasses all learning endeavors over the lifespan. The Internet, a large information database, readily available, may be used to support LLL. It is difficult to find the exact/correct information that one wants. Most of the search mechanisms on the WEB deal with isolated solicitations for information; being that the results for one determined request are identical, independent of the user or specific context of the user. It is possible improve the results using context information. The Learning Projects (LP) is a methodology based on the intentional and natural human activity of constructing projects. These LPs will give indications of the specific user's information necessities, providing the context of this information. Our system can assist the process of constructing LPs into Virtual Learning Environments by recommending documents. The IR model is a basic component because documents will be retrieved, classified based on the user's context, downloaded from the Internet and recommended to the user by artificial agents. The system builds domains ontology of the project using the user's interactions with LPs and concept maps as a strategy to build context.

Index Terms - Information Retrieval in Context, Intelligent Agents, Lifelong Learning, Learning Projects, Ontology, Virtual Learning Environment.

INTRODUCTION

Modern life activities demand the manipulation of large amounts of information and skills for decision making, so a constant learning process is needed throughout life. The concept of Lifelong Learning (LLL) as an educational strategy was defined as a response to this.

In the past years the Internet has become a large information database with billions of documents hosted in millions of sites distributed around almost all of the countries of the world. This entire library of documents is readily available, and may be used to support LLL. It is evident that not only the documents could be useful, but also all other multimedia tools that it provides, among these being: Dictionaries, Encyclopedias, Translation Tools, Multimedia Libraries and Learning Objects.

However, many times it is difficult to find the exact/correct information that one wants, due to the great volume of available documents on the Internet. Most of the search mechanisms on the WEB deal with the solicitations for information in an isolated way; being that the results for one determined request are identical, independent of the user or specific context of the user [1]. When we say that an Information Retrieval (IR) mechanism will work in a contextual way, it means that the documents will be retrieved considering cognitive and social factors, among others, related to the task, objective and intentions of the person. Cool and Spink [2] affirm that such factors compose the context of a user, although do not have an unanimous concept of this.

The Learning Projects (LP) is a learning methodology based on the constructive, symbolic, intentional and natural human activity of constructing projects. These LPs will give indications of the specific user's information necessities, and actual work tasks, providing the context of this information.

Another resource used in our paper are conceptual maps. They serve as a tool that concludes the learning processes. To the retrieval system, the maps are helpful because they make it possible for the apprentice to create models of the concepts he/she has been working with. Finally, the system agents that use anthologies also take advantage of the maps in order to offer the most accurate context.

We aim to develop a system to assist the process of constructing LPs into Virtual Learning Environments by recommending documents. The target users are individuals who are interested in using LLL autonomously, improving their education throughout life by using the Internet. In this scenario the IR model is a basic component because documents will be retrieved, classified based on the user's context, downloaded from the Internet and recommended to the user by artificial agents.

Computer Science advances make it possible for learning environments to assume a distributed architecture. The agent oriented paradigm emerged due to these advances, and showed sufficiently adequate in the development of this distributed learning environment, as seen in [4].

This present work uses intelligent agents as an approach to developing an environment that supports the process of constructing LPs using IR in Context. The choice of this approach, based on agents, was made because it is possible to develop more robust, modular, collaborative educational systems with these, where the tasks are distributed between some cognitive agents (knowledge and reasoning).

In section 2 we explain how LP may work as LLL. Strategy, in section 3 the Context Generation is detailed, the
information retrieval process and agents architecture are specified in section 4, section 5 presents the prototype of the system, and finally, in section 6, the final remarks are presented, plus, the current stage of development and future projects.

**LP as LLL Strategy**

According to Piaget [6], in his theory of the Genetic Epistemology, the phenomenon of learning depends on a constructive process that occurs through constructions and reconstructions of significant and logical systems of each individual. For this to happen, it is necessary that this individual can interact with objects (following Piaget this means: nature, physical world, culture, arts, sciences, languages...), with other people (society, institutions...) and now with technologies. We understand that learning occurs in a more efficient form when people involve themselves with the resolution to their problems, and their aim is to understand the roots of these problems that populate their universe. Papert [7] suggests that a good strategy for learning to occur is the construction of artifacts.

The Learning Projects’ development is based on the principles presented above. Where the learner starts its interactions with the environment, the lack of explanations for the physical, social and cultural phenomena of its world, start with an “inquiry question”.

The individual dedicates himself to make an inventory of the associated ideas with this inquiry question. These ideas can be grouped into two blocks: certainties and questions. These will be part of the methodological support for the trip that will be made. The questions are temporary, and during the process we will search for elements to clarify them. The certainties are provisory because they could be validated or refuted later.

To make this trip the learners make plans, organize the certainties and questions in a logical manner and choose an order for the inquiry. For each stage of the process they collect data, analyze, make connections, synthesize and create new questions and certainties.

As an example of this application is a Learning Project about ants. The inquiry question is “How does the organization of an anthill function?” Some initial certainties and questions are as follows:

- **c1**: There exists a subdivision of tasks in an anthill.
- **c2**: The ants work to protect the queen.
- **c3**: The ants responsible for searching for food and maintaining an anthill are the laborers.
- **c4**: The ants can communicate with one another.
- **c5**: There exists a hierarchy in an anthill.
- **c6**: An anthill has compartments with well defined purposes.
- **c7**: One of the compartments of an anthill is the dump.
- **c8**: The dump is the deepest compartment of an anthill.
- **q1**: How many ants are in an anthill?
- **q2**: How is the queen chosen?
- **q3**: How many queens are there in an anthill?
- **q4**: How do the ants communicate?
- **q5**: Can the ants adapt, learn, and behave in a different form when confronted with new situations?
- **q6**: Can misunderstandings (fights) occur in an anthill?
- **q7**: What intelligent activities have already been identified in anthills?
- **q8**: Does there exist a special compartment for food storage, or do the ants release it any place?
- **q9**: How is the queen chosen?
- **q10**: How many queens are in an anthill?
- **q11**: Does there exist a special compartment for food storage, or do the ants release it any place?
- **q12**: How many ants are in an anthill?
- **q13**: Can the ants adapt, learn, and behave in a different form when confronted with new situations?
- **q14**: Can misunderstandings (fights) occur in an anthill?
- **q15**: What intelligent activities have already been identified in anthills?
- **q16**: Does there exist a special compartment for food storage, or do the ants release it any place?
- **q17**: How is the queen chosen?
The LPP applied to Virtual Learning Environments has been used since 2002 by researchers at the Federal University of Rio Grande do Sul (UFRGS). The AMADIS project [5] is an example of this application, already being used in public schools.

**CONTEXT GENERATION**

A clear example of notion of context is presented by Ingwersen [3]. He states that a given word owes its context to the sentence to which it belongs; the sentence obtains its context from the paragraph it is part of, and finally, the paragraph gets its meaning from the text.

Our system operates in a similar manner organizing the items that it manipulates (concepts, certainties, doubts, notes, etc.) in a way that one item provides context to the other, which helps the system agents when they are asked to retrieve information.

The process is initiated by mapping the data that the user has been manipulating in the system’s interface. The creation of ontologies from concept maps has been cited in some papers [12] and [13]. These ontology maps all the elements of a learning project as well as its connections as exposed in Figure 2. The anthology model employed here is still simple since it does not make use of all its semantic power, although it is enough to deal with this specific case. Then it is shown the ontology created to the system by using the Protégé [15] tool and the Ontology Bean Generator plug-in for JADE [8]. The former serves to generate ontologies compatible with JADE’s agent’s library in a class format using Java programming language.

**INFORMATION RETRIEVAL AND AGENTS**

The agent’s system proposed in this paper is made by two agents; one for restoring and classifying information (Information Retrieval Agent) and another one for generating context and managing system’s ontology (Context Provider Agent). The function and characteristics of each one of the agents are detailed next:

- **Context Provider Agent (CPA):** receives the searches asked by IRA and adds terms to assist contextualization;
- **Information Retrieval Agent (IRA):** generates queries and submit them to CPA, runs the search and classify the results in accordance with the Learning Projects parts.

One of the first aspects to be considered is how the search terms are created to a given user and his/her projects. The answer is that to each learning project is generated one index of terms and each one of them has its relevance to the search calculated by tf-idf formula, which is one of the most popular classifying document skeins (1), (2) and (3).

\[
\begin{align*}
  w_y &= tf_y \times idf_y \\
  tf_y &= \frac{freq_y}{\max_x freq_x}
\end{align*}
\]
Where $freq_i$ is the number of occurrences of term $t_i$ in document $d_j$, $N$ is number of documents in collection, and $n_i$ is the document frequency for term $t_i$ in the whole document collection.

It was necessary to adapt the skein to better fit this context. A given document is mapped in one of the learning projects elements with which the user is interacting at the moment (concepts, certainties, questions, notes etc.) and the collection of documents is mapped in a group of elements of a learning project.

It is generated a search term index from a learning project and IRA can choose $n$ terms to research. From this list, IRA submits $n$ terms to CPA expands each one of the searches using as a basis the ontology that CPA already knows about that learning project.

CPA uses a technique known as Context Network Graph (CNG) that can serve for both classifying as well as expanding queries and that was chosen to carry out this task for being satisfactorily efficient, besides being simpler and faster than other similar techniques.

CNG is based on creating of a bipartite graph of term and document nodes, connecting a term node to a document node. In this model, every term is connected to all of the documents in which the term appears, and every document has a link to each term contained in that document. The weighted frequency values in the TDM correspond to weights placed on the edges of the graph. This construct is named Contextual Network Graph.

Classification obtained from certainties and questions of anthill learning project for query string “anthill queen” using the CNG technique:

- score: 1.1362722231413867 belongs to q3
- score: 0.5305469288332912 belongs to q5
- score: 0.512281932447919 belongs to q1
- score: 0.4472135954999579 belongs to q2
- score: 0.43956348224313135 belongs to q1
- score: 0.37200846792814624 belongs to c3
- score: 0.3511751345948129 belongs to c6
- score: 0.3124644028113322 belongs to q4
- score: 0.3069401310833123 belongs to q7
- score: 0.2992497704588229 belongs to q9
- score: 0.2964876345948129 belongs to c1
- score: 0.2886751345948129 belongs to c7
- score: 0.16193224092536315 belongs to q11
- score: 0.12752750715749256 belongs to c4
- score: 0.08333333333333333 belongs to q10

IRA is responsible for submitting the processed consults.

The searching mechanism is meta-search-based and uses search engines such as Google and Yahoo (both of then offer API’s connections in their document base). Finally, IRA classifies the results within the learning project.

In this classifying process it is used the same CNG technique, but now the purpose is to find out where the documents fit better and then recommend them by sending messages to the user.

### THE IMPARI PROTOTYPE

Our application was named IMPARI, acronym which means in Portuguese (Interface de Manipulação de Projetos de Aprendizagem e Recuperação de Informações / Learning Project Manipulation Interface and Information Retrieval). This prototype was completely elaborated using Java program language and free third part libraries to promote integration.

The Web interface was developed from Google Web Tool Kit [14] (GWT), which serves to construct Web 2.0 applications. Since it is a prototype we took advantage of an example that accompanies GWT Kit and we modified it as shown in Figure 3.

![Screenshot of WEB Interface Menu](image)

The agents are implemented with JADE (Java Agent Development Framework) (JADE, 2006). JADE is a software framework, fully coded in Java language, which simplifies the development of multi-agent systems through a middle-ware that complies with the FIPA specifications. This framework was chosen because it's actively updated, and it has a lot of facilities like GUI tools, debugging support, agents’ distribution across machines, agent common behaviors, communication handlers, etc.

### CONCLUSIONS AND FUTURE WORKS
The use of Virtual Learning Environments for promoting the LLL must be the reply of a vital problem for people who reach the height of their freedom. For this reason, the effort for the universalizing of the access to the Information and Communication Technologies, the Internet must be expropriated by everyone who believes in the technological development as a patrimony for all.

In this direction we aim the construction of a system to help individuals who want to keep themselves up-to-date in a continuous and autonomous form, considering the interactions and work tasks of the user. As part of this effort, we have developed a multi-agent system and a simplified Virtual Learning Environment that supports the LP Pedagogy. The conceptual modeling of the proposal is concluded and we are testing and debugging the prototype. As our next step, our objective is to finalize the implementation of the current system and then validate this system using our research group experience.

REFERENCES


AUTHOR INFORMATION

Luís Carlos Costa Fonseca, Federal University of Maranhão (UFMA), lccf@deinf.ufma.br.

Crediné Menezes, Federal University of Espírito Santo (UFES), credine@inf.ufes.br.

Rosa Maria Vicari, Federal University of Rio Grande do Sul (UFRGS), rosa@inf.ufrgs.br.

Jônatas Barbosa Soares, Federal University of Maranhão (UFMA), jonatas.soares@globo.com.