# An Intelligent Pedagogical Agent in CALL

Ognian Kalaydjiev

Bulgarian Academy of Sciences, CLPP, Linguistic Modeling Dept.,
25A Acad. G. Bonchev Str., 1113 Sofia, Bulgaria, ogi@lml.bas.bg

**Abstract.** The contribution presents a planning agent in an adaptive Web-environment for second language terminology learning. This pedagogical agent provides active or passive sequencing and suggests system-learner dialogue by open learner model. It performs as well task-specific personalized information retrieval of relevant readings.
**Thesis synopsis.** Thesis topic is "Intelligent Agents in Natural Language Processing Applications". The idea is to design and implement agents in information extraction, retrieval and filtering, and to apply them in a number of projects dealing with language technologies.

## 1 Completed work

The reported work in the area of intelligent tutoring and student modeling is completed after the first year of the PhD study, in the context of a CALL project. The research includes issues of both designing effective tutorial strategies and planning the content to be tailored to the individual user. It includes as well the choice of relevant texts to be displayed to the learner at a particular learning situation, as personalized information retrieval and filtering. The work done so far is related to the joint R&D project LARFLAST (Learning Foreign Language Scientific Terminology), funded by EC Copernicus programme [1].

**LARFLAST paradigm.** Larflast aims at the development of an intelligent, knowledge-based, adaptive www-learning environment for computer-aided learning of foreign language terminology. The prototype operates with about 200 terms and is tuned to the domain of financial English. The target users are adults non-native English speakers. The prototype integrates complex components like: *Open Learner Model* (OLM, developed by CBLU, Leeds), the natural language understanding system *Parasite* (developed by the Manchester team, UMIST) and *Web-agents - spiders* for searching texts in Internet, developed in Montpellier (LIRMM). In this way the planning agent is supposed to balance and integrate the modules dealing with intelligent tutoring, student modeling, and web-based text retrieval in the context of language learning.

Domain knowledge is kept as a KB of conceptual graphs. The Learner Model (LM) stores clauses reflecting the learner moves and her performance in drills. Drills are especially designed to test domain concepts and KB facts and the drills' goal is explicitly recorded in drill annotation. Thus the LM records can be directly targeted to KB units (not to linguistic ones). LM contains four kinds of clauses: *(i)* **know** - the learner knows a domain concept (fact); *(ii)* **not_know**

- the learner doesn't know a domain concept (fact); *(iii)* **self_not_know** - the learner has registered himself that he doesn't know a domain concept (fact); *(iv)* **know_wrongly** - the learner has built knowledge (concept/fact) that is considered as wrong by the system (eventually, might need to be corrected). Adaptivity of the presentation (next drills and tutoring materials) is provided by a pedagogical agent (PA), implemented by the author (see [1]).

**The planning agent.** PA plans future learner's moves between *(i)* performing drills (active sequencing), *(ii)* suggestion of readings (passive sequencing) and *(iii)* suggestion of dialogue by OLM. At present the planning is reactive and local [2]. Since considerations concern presentational as well as educational issues, according to the terminology in [3] we would classify the planner as performing some aspects of instructional as well as content planning. PA has two main strategies for active sequencing - *local* and *global*. The local strategy plans moves between drills testing different characteristics of one concept. Its main goal is to create a complete view about learner's knowledge about this concept. This strategy chooses drills with increasing complexity when the learner answers correctly and gives again previously completed drills if the student has performed poorly. The global strategy plans movements between drills testing different concepts, according to their place in the financial ontology. PA chooses next learner's movement depending on: *(i)* the predefined drill's goals, *(ii)* KB items, *(iii)* concept weights defined in the drills' annotations and *(iv)* current learner's score.

Below we focus on the recently implemented modules providing personalized information retrieval and information filtering in Larflast. This task is not considered in detail in [1] and therefore represents an original contribution.

The overview [4] states that the currently available www-ITS systems are most often sets of static, hyperlinked html-pages. Larflast aims at a more elaborated, intelligent decision concerning personalized presentation of tutoring materials. The idea is inspired by the Web-context, where many financial sites expose and frequently update texts. To show to our learner readings containing most relevant information, we have to support and contiguously update a database of documents. (The collection itself is performed by Web-agents). Our planer thus operates with: *(i)* a data base, containing financial texts collected from Internet and *(ii)* a relevance measure, showing for each text the percentage of its relevance to the domain terms $T_1$, $T_2$, ..., $T_k$. These terms are juxtaposed to KB concepts. The relevance measure is associated automatically to each text by a LSA-module [5] (an original implementation of Sofia team).

The goal of PA is to select which text is most relevant to be displayed as a tutoring material (reading) at the particular learning situation. At each situation, LM keeps track of the terms which are unknown or wrongly known to the learner. The text with higher relevance to all these terms has to be selected. Most generally, this is done as follows: The learning situation is estimated with respect to the terms $T_{n1}$, $T_{n2}$, ..., $T_{nm}$ which appear in **not_know**, **self_not_know** and **know_wrongly** LM clauses. Actually we operate with the KB concepts, juxtaposed to these terms. The estimation is unique for the current learning situation, it is calculated for each term $T_{ni}$ from $T_{n1}$, $T_{n2}$, ..., $T_{nm}$ and represents the sum of: *(i)* the predefined weight of the corresponding concept in the KB hierarchy, an integer between 1 and 10; and *(ii)* closeness of the focused concept $T_{ni}$ to the concepts $T_{n1}$, ..., $T_{ni-1}$, $T_{ni+1}$, ..., $T_{nm}$. All pairs ($T_{ni}$, $T_{n1}$), ..., ($T_{ni}$, $T_{ni-1}$), ($T_{ni}$, $T_{ni+1}$), ..., ($T_{ni}$, $T_{nm}$) are considered and the values "close-distant" (respectively "1-0") are summed. Two concepts are *close* if they are either linked as child-parent in the KB hierarchy, or they are sisters according to the same partitioning perspective. Otherwise the concepts are considered as *distant* ones.

After calculating the sums $S_1$, $S_2$, ..., $S_m$ for the terms $T_{n1}$, $T_{n2}$, ..., $T_{nm}$, the integers $S_1$, $S_2$, ..., $S_m$ are sorted in decreasing order. In a sense, the terms in question are "sorted" in decreasing order according to their relevance to the particular learning situation. Let $T_{r1}$, $T_{r2}$, .., $T_{rm}$ be the new order (by relevance). For each term $T_{r1}$, $T_{r2}$, .., $T_{rm}$ the planner finds the set of relevant texts, available at the moment in the text data base. Starting from $T_{r1}$ to $T_{rm}$, the planner looks for texts maximally relevant to all terms. In this way PA proposes readings that provide "maximal relevance" to the unknown terms, taking into consideration the estimation of terms' weights. In other words, the idea is to select readings giving preference to: *(i)* term importance in the domain and *(ii)* term closeness (in order to explain in one document as many terms as possible). Since Larflast project is entering the final evaluation phase, the planner and its strategy for choosing relevant text will be soon evaluated too. Therefore small modifications of the above-described heuristics might be expected.

## 2 Future work

Currently Larflast KB encodes 200 terms. Experiments in personalized text retrieval include more than 300 texts. The future work concerns the elaboration and mainly assessment of the intelligent planning agent, with focus on the evaluation of the user-tailored presentation.

## References


1. Sv. Boytcheva, O. Kalaydjiev, A. Nenkova and G. Angelova. Integration of Resources and Components in a Knowledge-Based Web-Environment for Terminology Learning. In Proc. AIMSA-2000, Springer, LNAI 1904, pp. 210 - 220.
2. McCalla, G. The Fragmentation of Culture, Learning, Teaching and Technology: Implications for the AI in Education Research Agenda in 2010. In IJAIE (2000), 11, pp. 177-196.
3. Vassileva, J. and B. Wasson. Instructional Planning Approaches: from Tutoring towards Free Learning. In Proc. EuroAIED'96, Lisbon, Portugal, 1996, pp. 1-8.
4. Brusilovsky, P. Adaptive and Intelligent Technologies for Web-based Education. In: C. Rollinger and C. Peylo (eds.) KI 1999 (4), pp. 19-25.
5. Deerwester S., Dumais S., Furnas G., Laundauer, T. and Harshman R.: Indexing by Latent Semantic Analysis. JASIS 41(1990), pp. 391-475. LSA URL: http://lsa.colorado.edu (1990-99).