



HAL
open science

ISNLP on AI for Education - Exploratory action 2019: Exploratory experiments combining Information Systems and Natural Language Processing

Annie Foret

► **To cite this version:**

Annie Foret. ISNLP on AI for Education - Exploratory action 2019: Exploratory experiments combining Information Systems and Natural Language Processing. Cominlabs days, Nov 2019, Rennes, France. hal-02462991

HAL Id: hal-02462991

<https://inria.hal.science/hal-02462991>

Submitted on 31 Jan 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



ISNLP on AI for Education

Exploratory action 2019

Exploratory experiments combining Information Systems and Natural Language Processing

- Teams
- SemLIS – IRISA (resp.)
 - TALN - LS2N (resp.)
 - Druid – IRISA
 - Hycomes - IRISA
 - LogicA – IRISA
 - with: Arènes – Rennes 1
 - ISTIC – Rennes 1
 - LIG – Grenoble
 - LOUSTIC – Rennes 2

Objectives

We proposed experiments to enrich texts (in various formats) that serve as pedagogical support.

This comes in two major ways:

- extraction of major parts from full text, automatic or interactive.
- completion of lacunar texts (eg. slides or program comments)

To help learners through enriched pedagogical documents / texts in comp. sc., languages ... Information NLP

Major terms are to be connected with referring or enlarged information (terminology, definition, formula, symbols, etc.).

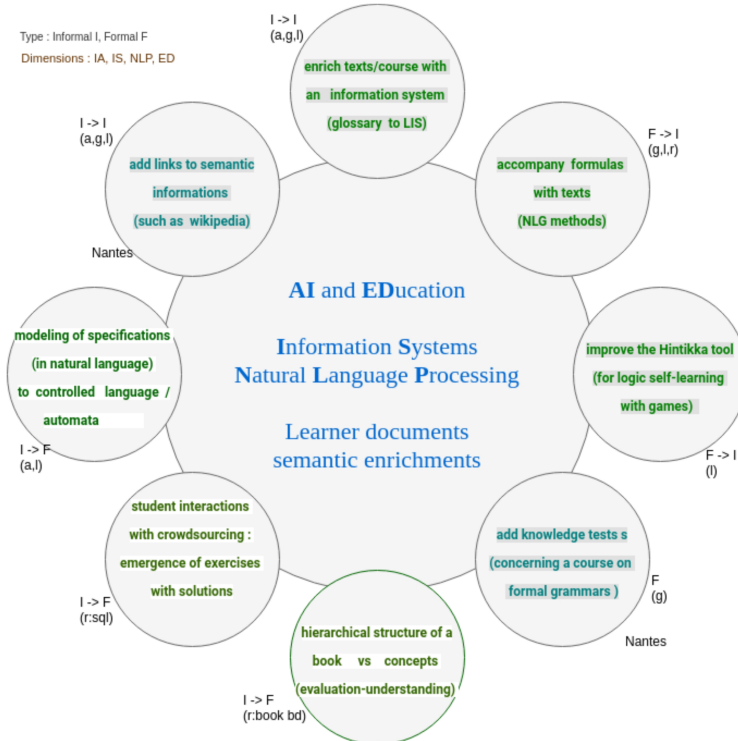
The targeted complexity may go from a simple glossary to a rich semantic network and may include interactive technology.

Learning context: between courses, before exams,... self-learning

Texts/concepts: enrichments, avoid confusions, ... self-correction via semantics + new exercises, precise terminology, explanations

These features are not topic-specific (science, LV1, ...)

Experiments overview



Case study : courses on automata (a), grammars (g), logics (l), query languages (r) → level : L1-L2-L3
 Developpement : in python [py], java [ja], angular js [js], php [ph], ... → integration, shell linux version, web version
 proof-of-concepts on gitlab, vim, ...

Learners

The experiments considered students in computer science, in particular: their learning of formal topics and of query languages.

Pitfalls examples: definitions-conventions and variability

partial vs (not) total (for a function, for an order)
 N including 0 or not
 antisymmetric vs (not) symmetric (for a relation, for a matrix)
 C w.r.t. (not) =
 ...
 + ill Formal-Informal renderings: $\forall x(\wedge .)$ instead of $\forall x(. \implies .)$

For Course Companions in computer science

Questionary analysis by LOUSTIC

Récapitulatif des recommandations	
Habitudes des utilisateurs	Se baser sur la lecture et prioriser les ressources dispensées en cours et mises en ligne par les enseignants
	Ne pas contraindre l'outil au travail de groupe
	Anticiper les périodes d'examens (utilisation plus importante)
Difficultés des utilisateurs	Cible à privilégier : étudiants en Licence, travailleurs et en difficultés
	Soulager la charge de travail / Ne pas ajouter de charge de travail
	Pallier le manque de motivation (gamification, contenu ludique, ...)
	Concrétiser les cours (exemples, exercices, ...)
Attentes des utilisateurs	Aider la compréhension (supports des cours, objectifs clairs, ...)
	Faciliter l'interaction profs-élèves (espace de discussion / forum / FAQ)
	Synthèse des points importants, définitions & formules
	Proposer des exercices, annales & exemples (et leurs corrigés)
	Regrouper l'ensemble des ressources des cours et un maximum de ressources externes
Remarques des utilisateurs	Proposer un suivi (évaluation de l'apprentissage) avec système de progression (gamification), tout en restant ludique
	Espace de discussion profs-élèves et/ou élève-élève
	Faciliter l'organisation temporelle (indicateur temps passé, planning)
	L'utilité, l'accessibilité, la praticité, l'efficacité et la facilité d'utilisation doivent primer sur la stimulation, la créativité et l'esthétisme
	Accessible via ordinateur et smartphone (en ligne avec hors ligne possible)
	Compatibilité avec les supports de cours (slide)
Outil activable / désactivable	
Favoriser l'apprentissage actif (exercices/tests)	
Accessibilité, flexibilité et adaptabilité	
Proposer des exercices (et leur corrigé) et des quiz offrant un suivi dans l'apprentissage	
Outil ergonomique et esthétique	
Possibilité d'évaluer l'outil (amélioration continue)	

Tableau 21 : Récapitulatif des recommandations

Informal to Informal

Workflow on Terminology

SemLIS

The new workflow goes from texts or pdf to a selected terminology. It has been tested on a course at L2 level and can be run on a terminal or as a web interface. It can deal with multiword expressions (such as "diagramme de Venn") and symbols (such as set operations) in a pdf file.

This work prepares the links to definitions and concepts for relevant terms/symbols.

Enriching a lacunar text with external links from Wikipedia

TALN

We aimed at identifying the concepts related to computer science, using Wikipedia data (in French). Categories and portals are useful sources but do not lead easily to a clean list. The new tool identifies the words of a text that correspond to the domain (computer science) and proposes to add URL links to Wikipedia pages. It is currently semi-automatic.

Wikipedia ex. éléments de E sont toujours comparables, c'est-à-dire que \leq est une relation d'ordre total, et que l'ensemble E est totalement ordonné par cette relation. Une relation d'ordre sur E est dite partielle si elle n'est pas totale, et E est alors partiellement ordonné. Il est à noter qu'en anglais, on désigne par *ordre partiel* un ordre quelconque, qui peut donc

Formal to Informal via NLG, games, exercises

Proof of concept involving Natural Language Generation (NLG)

SemLIS

The new prototype goes from formulas to texts in both languages french and english. The internship was asked to use SimpleNLG. He developed the prototype for two test cases (taught at L2 and I3 level): a fragment of logic and a fragment of XPATH.

Crowdsourcing of a tutorial document by students, from students DRUID

Using the HEADWORK platform, our internship developed a step-by-step form where participants can suggest new questions for the tutorial, and propose a correction. By majority voting, they collegially decide which question is interesting and which is the best correction. In case of tie in majority voting, a call to an oracle teacher is performed. The final tutorial document is automatically generated with its correction.

Limited time responses, fancy, interactive HTML web pages were also obtained. Every step is performed using automata driven by queries on a database.



Improving a tool for learning modal logic

LogicA

Hintikka's World can be used for learning modal (dynamic epistemic) logic with examples. Epistemic Situations are described by Kripke models. For card games the number of possible worlds is gigantic. This is why symbolic models based on BDDs are used. Our internship has proposed a new software architecture, for better GUI performances. He used the asynchronous programming paradigm.

Learning formal grammars and the writing of syntax analyser

TALN

The created educational software proposes to study formal grammars and the writing of syntax analyser by hand. From a given formal grammar (generated randomly), the learner must answer several questions that lead to a final program (written in C) that can analyse this grammar.

The questionnaire shows a demand for more training exercises.

Informal to Formal

Evolution of concepts via term co-occurrence graphs

DRUID

We worked on enriching learning resources to analyze and to understand the evolution of involved concepts: in particular, in scientific articles, from a large corpus of scientific publication (arXiv) [workshop LEG'2019]. This work is based on the analysis of the term co-occurrence graphs, where we applied spectral techniques.

In a similar way, we could extract terms from two widely used textbooks (Deep learning and Android programming) in PDF format.

Annotated Corpus for Modelling

SemLIS

With the aim of assisting the engineer to design a system, the evaluation and validation of a processing chain around system design remains a challenge. We thus defined study cases and started an annotated corpora to prepare a semantic extraction from specifications.