Search Engines in Learning Contexts: A Literature Review

https://doi.org/10.3991/ijet.v17i02.26217

Jaurès S. H. Kameni[1](✉), Bernabé Batchakui[1], Roger Nkambou[2]
[1] Department of Computer Engineering, National Advanced School of Engineering, Yaoundé, Cameroon
[2] Department of Computer Science, UQAM, Montréal, Quebec
jskameni@gmail.com

Abstract—The web is one of the primary sources of information for finding learning oriented documents. In addition, the main suitable way to find information and documents on the Internet is by using search engines. Search engines are constantly improving in terms of selection algorithms and in terms of the Human Machine interface (HMI). Also, these search engines are the basis of a new field of research called Search-As-Learning. The Search-As-Learning explores information search environments to enhance learning during user search tasks. This work focuses on our view of the state of the art in the field of Search Engines in learning context and Search-As-Learning, stressing on the most recent research. We conclude by highlighting the current shortcomings on improvement of the learning aspect within search engines, and present next work which will be the association of a layer above the traditional search engines to promote the appropriation of content during search task for a learning context.

Keywords—information retrieval, search engine, Search Engine Result Page (SERP), Search-AS-Learning, learning

1 Introduction

Search engines are the most popular way to find information today. Even more, the use of a search engine to perform a web search is the most popular activity on the Internet [1]. In the past, documentary search was carried out mainly in libraries. Nowadays, searching for documents via the Internet is an essential practice for Internet users, particularly for teachers and learners. In this momentum, search engines have evolved into a refined search and sophisticated display of search results [2].

In terms of interface, it can be said that any search engine displays the return of queries using a page called Search Engine Result Page (SERP) [3]. Researches on studying the layout of search results pages are ongoing. Early results on this have shown that the layout of the results influences user selection [4]. In the literature, internet queries are mainly categorized into three classes: navigational queries, informational queries and transactional queries [5]. From the year 2007, new results have

shown that the representation in enriched snippet [6], [7] on the result page was essential for informational queries [6], [8]. Today, in addition to traditional snippets, search engines are full of rich snippets [9], for example: home pages, events, products, job offers, Images, Tweets, Advertising, Videos, Addresses, Maps, News, Knowledge Graph.

eLearning [10] is extremely important in education nowadays. We underline for example that this method has been serving as the main relay during the period of expansion of covid19 [11]. Just recently, we have the advent of Search-As-Learning (SAL) which is a subarea of eLearning, surrounding learning by doing information search tasks on a system like Google, Bing, Yahoo Search, etc. Thus, researchers recognize search systems as rich online spaces in which users can learn and discover new knowledge while interacting with online content [12]. Around SAL, new research has focused on the acquisition of knowledge during information retrieval tasks on search engines [13], [14], [15], [16], [17]. In addition, it has been shown that SAL mainly relates to informational search task [13], [15], [16]. However, for a search task which is oriented learning as example, SERP produces relevant snippets, but also, irrelevant excerpt related to advertising, products, events and the like. Thus, one consequence is that these non-educational snippets grab the attention [18] and clicks of learners to produce cognitive overload. Therefore, with SAL, a new challenge in the field of information retrieval is to optimize results with regard to educational objectives [19], [20].

This work is mainly aimed to present a literature review around search engines in learning contexts. We first present some bases on information retrieval in traditional search engines, then we follow with the state of the art about Search-As-Learning. Finally, we conclude by highlighting the limitations in terms of the search result page in a knowledge acquisition context.

2 State of the art on search engines in learning contexts

This section presents related works that examines the different web strategies (algorithmic and human-machine interface) used in setting up SERP, and then present the state of the art in the subdomain of Search-As-Learning. Particular attention will be paid to the limitations of existing approaches to SERP oriented for learning.

2.1 Information retrieval and search engines

After significant advancements in search engine selection and ranking algorithms, researchers looked at various aspects of SERP and explained how their design affects user behavior. Here we provide a summary of the various aspects that have been investigated. We first focus on the selection and ranking of results in search engines in general, then we study: the layout of the SERP; the size of the SERP; how the text of the snippet is generated; and how much text should be presented in each summary of results. Some work on these has been done by David Maxwell et al [8].

Bases for selection and ranking of results in search engines

a. Information Retrieval Models

In order to find the information that corresponds to a user need, classical Information Retrieval Support System (IRSS) are used. They are based on an architecture where the user is just a consumer of the content. In general, an Information Retrieval (IR) model is characterized by the document and query representation model "F" as well as the document-query matching process "RSV (q, d)". It was formally defined by a quadruple (D, Q, F, RSV (q, d)) [2] where:

- D is a set of document representations,
- Q is a set of queries,
- F is a framework for modeling document representations, queries and their relationships,
- R associates a real number to document-query-pairs (ranking),
- RSV (q, d) is a ranking function which associates a real number with a query $q \in Q$ and a document representation $d \in D$. Such ranking defines an ordering among the documents with regard to the query q.

There is a wide variety of models in the literature [2]. Figure 1 presents a classification of the most important, mainly distributed around three families: Boolean, vector and probabilistic models.

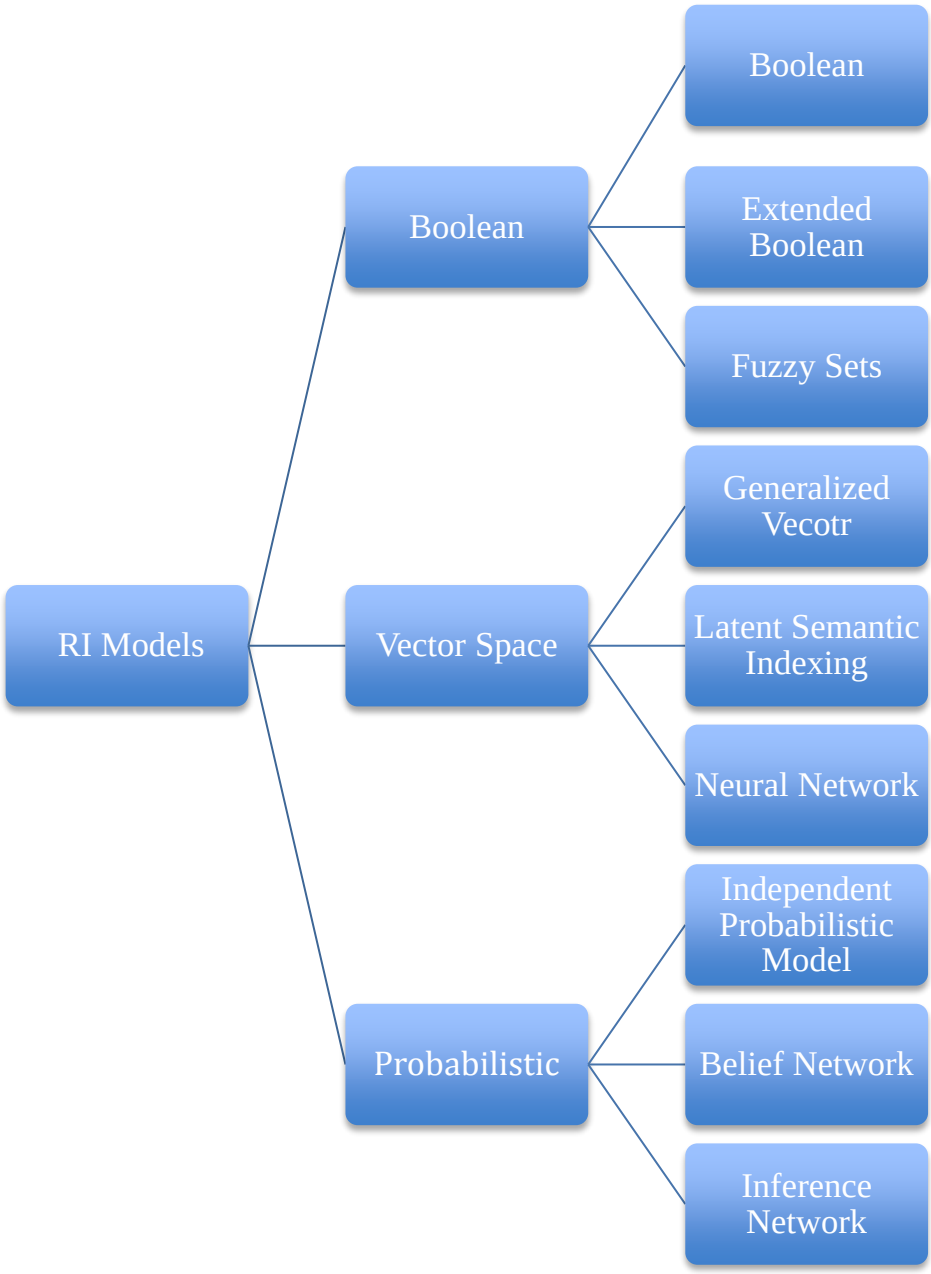

Fig. 1. Classification of Information Retrieval Models

b. Working principle of search engines

Search engines as we know today, although distinct in their ranking methods, all operate in quite similar ways. Their operation can be broken down into four stages [21], [22]:

- Step 1: Information crawling;
- Step 2: Indexing the pages;
- Step 3: Processing requests;
- Step 4: Ranking of the results.

c. Ranking of the results

The main quality of a search engine is the efficiency of its ranking system [23]. The main filing systems are as follows:

– Sorting by relevance: the relevance sorting method is used by most search engines [24]. Sorting by relevance allows you to evaluate the ranking of a page according to five factors: the weight of a word in a document; the density of a word; the weight of a word in the database, the correspondence; the close relationship.
– Sorting by popularity: sorting by popularity favors web pages with a large number of backlinks. This sorting method was used by Google in 1998 to create the PageRank algorithm [25].
– Sorting by audience: sorting by audience measurement favors web pages which have the highest "quantity" and "quality" of visits.

Restitution of search engine results

a. Search engine result pages and their presentation

Amongst the first works on SERP is that of Dumais and Chen on the presentation of snippets of results [26], [27]. This work investigated approaches to automatically classify result summaries for users, similar to the categorization approach used by early search engines. Subjects in a user study found the interface to be easier to use than the traditional "ten blue links" approach. Indeed, they were 50% faster to find the information displayed in the categories. This work was then extended by Dumais et al [27], where they explored the use of hover text to present additional details on search results based on user interaction. The search was found to be slower with the hover, possibly due to the fact that explicit decisions about when to (respectively not to) search for additional information where needed (respectively not needed).

Arrangements based on a grid have also been tested [28], [29], [30]. For example, Krammerer and Beinhaur [29] examined the differences in user behavior when interacting with a standard list interface compared to a tabular interface, as well as a grid-based layout (summary results placed in three columns). Grid users spent more time reviewing the summary results. The approach has shown promising results in overcoming problems such as positional bias [31], as observed by Joachims et al [32].

Marcos et al [9] performed an eye-tracking study, examining the effect of user behavior during their interactions with the SERP, and determining whether the richness

of the results snippets provided on a SERP had an impact on the users' search experience. The result demonstrates that the representation of an enriched text document during the presentation of the results could have a positive effect on the evaluation of the relevance and the reformulation of the queries [33]. However, classification remained the most critical factor in the perception of relevance. The enriched summaries were also reviewed by Ali et al [34] in the context of navigation tasks. After this review, a positive relationship is shown between the length and informative nature of the abstracts and their attractiveness. However, we do note a loss of performance.

b. Generation of the snippet text

Results summaries let users know whether a document is likely to be relevant or not [35]. As a result, research was undertaken to examine different types of snippets and the length of an snippet. The work initially focused on how these extracts were to be generated [36], [37], [38], [39], [40]. This early work was proposed to summarize documents by query (query-oriented summaries) or keywords in context, as opposed to simply extracting representative sentences or main sentences from the document [41]. Tombros and Sanderson [39] showed that subjects in their study were likely to identify relevant documents more precisely when using summaries of extract text associated with a query, compared to summaries of simply extract text generated from the first sentences of a given document. Recently, text-based snippets have been shown to be preferred on mobile devices [42].

When constructing snippets using query-oriented summaries, Rose et al [43] found that snippets offered a perception of the quality of the result. If the snippet texts contained truncated sentences or many fragmented sentences, users perceived the quality of the results more negatively, regardless of their length. Kanungo and Orr [44] maintained that readability is a crucial presentation that must be taken into account when generating a query-oriented text summary. Clarke et al [7] found that results with very short (or missing) snippets containing fewer query terms are not quite readable, and have a reduced number of clicks.

Recent work has examined the generation of snippets from more complex angles ranging from manipulation of underlying indexes [45], [46] to language modeling [35], [47], as well as the use of search data from users to improve the generation process [48], [49]. Approaches from previous generations may also fail to consider which parts of a document are actually useful for users. Ageev et al [48] incorporated post-click searcher behavioral data, such as mouse cursor movements and scrolling through documents, into a new model, producing behavior-oriented snippets. The results showed a marked improvement over a strong base for generating snippet fragments. Temporal aspects were also taken into account. Svore et al [50] conducted a user survey, showing that they preferred snippet text with trending content in snippets when searching for trending queries, but not for general queries.

c. Number of results per page

Examining behavioral effects on mobile devices when interacting with SERP has recently been the subject of much research (e.g. [51], [52]), with each device showing

a different number of results. Recent research has shown that the value of RPP (Results per Page) can influence the behavior of document searchers [32], [53]. Understanding this behavior can help guide and educate those responsible for designing user interfaces.

In a Google report, Linden [54] indicated that users wanted more than 10 RPPs. Despite the fact that increasing RPP resulted in a 20% drop in traffic, it was speculated that this was due to the extra time required to send the SERP in addition. However, this drop in traffic can be attributed to other reasons. Oulasvirta et al [55] deals with the paradox of choice [56] in the context of search. This is because when there are more options for results, especially if those results are all very relevant, it can lead to poorer choice and worsen user satisfaction. Oulasvirta et al [55] found that presenting users with a list of search results with six items was associated with higher degrees of satisfaction, confidence in choices and a perception of caution than a list of 24 elements. Kelly and Azzopardi [57] broadly agreed with the findings of Oulasvirta et al [55]. The authors conducted an inter-subject study with three conditions, with subjects assigned to one of three interfaces: the base interface, indicating 10RPP (the "ten blue links"), and two interfaces displaying 3RPP and 6RPP respectively. The results showed that people using the 3RPP and 6RPP interfaces spent significantly more time reviewing the best results and were more likely to click on higher ranked documents than those on the 10RPP interface. The results also suggested that subjects using interfaces displaying less RPP found it comparatively easier to find relevant content than those using the 10RPP interface. However, no significant difference was found between the number of relevant elements found across interfaces. Currently, 10RPP is still considered the norm [58].

d. Snippet lengths

The lengths of the snippets were examined in different ways. A user study by Paek et al [59] compared a user's preferences and usability with respect to three different interfaces for displaying snippets of results. In the first case in their work, for information retrieval tasks, the interfaces displayed a normal SERP (i.e. a two-line fragment snippet for each summary, with a clickable link); an instant interface, where an expanded extract was displayed by clicking on it; and a dynamic interface, where hovering the cursor would trigger the expanded code snippet. Instant View has been shown to allow users to complete given search tasks in less time, with half of participants preferring this approach. The pioneering work of Cutrell and Guan [6] explored the effect of different lengths of snippets (short: 1 line, medium: 2-3 lines, and long: 6-7 lines). They found that longer snippets significantly improved the performance of information tasks (e.g., "Finding an airport address"). Users got better performance for information queries as the snippet length increased. This work was followed by Kaisser et al [60] who conducted two experiments that estimated the length of the result snippet based on the type of response (for example, finding a person, time, or location), and comparing the results of the selected snippet lengths to user snippet preferences to see if it could be predicted. The preferred extract length was shown to depend on the type of response expected.

Research work on examining which snippet sizes are appropriate for mobile devices is ongoing. This is important since the screen sizes are smaller. Also, snippet text considered acceptable on a computer screen may involve considerable scrolling or sliding on a smaller screen. Kim et al [61] found that subjects using longer snippets on mobile devices exhibited longer search times and similar search accuracy in informational tasks. Longer reading times and frequent scrolling or swiping (with more view movement) were exhibited. So the longer clips didn't look very useful on a small screen. Thus, an instantaneous or dynamic approach (according to Paek et al [59]) could also be useful for mobile search.

The presentation of summary results has an important effect on a user's ability to judge relevance [35]. Relevant documents may be overlooked due to non-informative summaries. Conversely, in informative excerpts, irrelevant documents may be reviewed due to a misleading summary. Recent work by David Maxwell et al [8] showed that longer summaries also lead to an increase in the cost of examinations. However, in the context of a search in a learning context (searches in a learning context are mainly related to information queries [13], [15], [16]), some works are favorable to the display of a lot of information around the subject, to allow the user to explore various aspects of the subject. Thus, it would therefore not be in vain in a learning process, to read an element that would seem irrelevant because it fits into the contours or the contexts of the subject.

Remarks on search engines and learning. Works on information retrieval via search engines presented previously does not particularly integrate the specific case of learning. However, as searching information in learning context is informational [13], [15], [16], we understand from previous research that a result page in a learning oriented search framework must have 10 results per page [58], summaries must be query based [42], [43] with an informative text (6-7 lines) as underlined in the last paragraph of Section 2.1.2.a and in Section 2.1.2.d. Note also that the new field called Search-As-Learning, is the specific field investigating the improvement of search engines in a context of education and learning.

2.2 Search-As-Learning

In the literature, studies in the context of "Search-As-Learning" can be divided into five main themes: 1-search tasks related to learning, 2-exploratory search oriented learning, 3-expertise and learning, 4-evaluation of learning during search task, and 5-optimization of search results for education.

The recent work of Ran Yu et al [15] presented a state of the art by grouping the work around Search-As-Learning in the following way: Detection of the intention of web search sessions from user interactions, understand the learning process and outcomes, and support learning while searching the web. Table 1 provides a correspondence between the literature around Search-As-Learning and the work of Ran Yu et al [15].

Table 1. Grouping research works around Search-As-Learning

	Consolidation of the work of Ran Yu et al. [15]	**In a more specific context of the Search-As-Learning literature**
1	Detection of the intention of web search sessions from user interactions	Search tasks related to learning
2	Understand the learning process and outcomes	-Exploratory search oriented learning -Expertise and learning -Evaluation of learning during search task
3	Support learning while searching the web	-Optimization of search results for education

Search tasks related to learning. Learning-related search tasks can be complex, requiring multiple queries and considerable time spent in searching and browsing. A study by Bailey et al [62] on how users engage in such tasks using search engines describes a taxonomy of web search tasks, based on 4 months of log data from Google, Yahoo, and Bing. This taxonomy included certain tasks related to learning, covering the exploration of subjects ("discover more information about a specific topic"), the search of fact diverse ("find facts about a person" and "find a specific miscellaneous fact") and procedural learning ("learn how to perform a task").

Subsequently, Eickhoff et al [63] analyzed the fraction of search sessions involving a procedural or declarative knowledge intention. They found after selection criteria in the log files that 3% were clearly learning work from procedural or declarative knowledge. Raman et al [64] also noted the large number of intrinsic diversity search tasks. In fact, the various intrinsic search tasks are exploratory web searches intended to explore and learn more about multiple aspects of a specific topic. Jansen et al [65] found that information retrieval is a learning process with unique search characteristics specific to particular learning levels of Bloom's taxonomy [66]. Other recent works have attempted to assess the motivation of users to perform information retrieval tasks. For example, Kim et al [67] characterized this motivation in terms of a user's willingness to search and browse documents which are well above his reading level.

Exploratory search oriented learning. Learning oriented search activities often involve multiple interactions in the search process and the processing of multiple sets of findings that need to be interpreted in depth by researchers. Marchionini [68] asserts that search activities that promote learning focus more on "the acquisition of knowledge, the understanding of concepts or skills, the interpretation of concepts, and the comparisons or aggregation of data / concepts". Therefore, search activities that support learning require human participation on a more continuous and exploratory way during the search process. Exploratory search, which emphasizes broader information-seeking strategies focusing on deeper understanding rather than rapid factual responses, has become an alternative paradigm for fostering learning and inquiry in research [68], [69]. White et al [69] propose learning outcomes as an important future assessment method for exploratory search [69]. The work of Ran Yu et al [15] provides an overview of the challenges and research approaches to detect, understand and support lifelong learning in web search.

Expertise and learning during search tasks. Expertise is a dynamic characteristic of users that reflects learning over time. Wildemuth [70] examined how domain expertise was reflected in users' choice of search strategies, concluding that domain novices tended to exhibit search strategies increasingly similar to those of more experienced users as they learned more about the subject. Previous work [71], [72] has characterized domain expertise and search behavior in terms of metrics that can be derived from search logs, generally focusing on long-term patterns of behavior from session to session.

In one of the first large-scale journal-based research studies to examine the functionality of session-level tasks where people explicitly seek new knowledge, Eickhoff et al [63] examined assignment changes for these expert measurements. They focused on two types of specific knowledge acquisition: procedural knowledge (how to do something) and declarative knowledge (knowing facts about something). The authors found evidence of both learning progress within a single session and the persistence of learning between sessions. A significant portion of the new query terms came from results page snippets and recently visited pages, showing that the search process itself has helped increase the user's domain knowledge. Other studies like Zhang et al [72] have attempted to predict domain knowledge from users' search behavior. Finally, N Bhattacharya et al [17] recently showed that the differences in participants' knowledge changing are reflected in their search behavior and visual measures on results pages and towards pages containing content. In his survey paper, Vakkari [73] described more in detail how learning occurs in the search process.

Evaluation of learning during search tasks. A few studies have attempted to identify indicators of learning during the search process. Vakkari et al [74] found that students' level of knowledge about their subject could predict the characteristics of search queries, in the sense that students who know less about the subject are likely to use fewer search terms more vague in their queries. In another study with medical students, Vakkari and Huuskonen [75] found that efforts in the search task did not lead to better search results, but improved research works on a subject. Several information seekers have designed search methods to study learning as a measure of search outcomes. In one of the previous studies measuring learning, Hersh et al [76] showed how search allowed students to answer more questions in a post-search questionnaire. Instead of a quiz, Kammerer et al [77] asked study participants to write a summary of the topic after using the MrTaggy exploratory search interface, to assess learning. The quality of the abstract was rated based on topic-specific criteria, including the number of reasonable topics, the overall quality of the topic description, and the number of arguments. Wilson et al [78] developed systematic techniques for measuring the depth of learning at three levels: quality of facts, interpretation of data in statements, and use of criticism. Still on the assessment of learning during search, recent work by Ujwal et al [13] has shown that users acquire important knowledge during search sessions on topics with which they are less familiar.

Optimization of search results for education

- **User interface and interaction:** The web today has several online learning platforms. Many Learning platforms are environments offering massive data and con-

tent for learning [79]. Indeed, learning by doing information search tasks is also held within these environments and they have interfaces of the following components for learning: forum, content, board, collaboration, messages which are important for learning and tailored to the profile of learners. The works around the MOOCs also allows us to have enough content and a search environment with an adequate interface taking into account the profile of the learner [80]. However, in general-purpose search engines, learning support is not taken into account also because of the versatile nature of these environments and the variety of tasks performed in them. A central question for research in this area is how interfaces can be adopted to improve learning performance, even in versatile search engine environments. An attempt aimed at improving user engagement in learning-oriented search tasks by providing a richer representation of retrieved web documents was made by Arora et al [81]. Specifically, they explored methods to find useful semantic concepts in retrieved documents, with the goal of creating improved document surrogates for presentation in the SERP. Kodama et al [82] studied the relationship between the Google mental models of college students and their information search skills. During the Kodama et al [82] experiment, the students showed limited understanding of how search engines work. Additionally, this research suggests that developers and interface designers make the inner workings of the engine more transparent to learners. Recently, Qiu et al. [83] produced a conversational interface for search in learning context which can improve user engagement, augment user long-term memorability, and alleviate user cognitive load.

- **Retrieval and ranking:** As current search engines are optimized by taking into account a need for information without considering the learning intention of a query, research on the optimization of extraction and ranking algorithms based on learning needs are relatively rare. In 2007, I. Bosnić et al [84] propose a system indexing learning resources with SMIL which helps finding the right data in multimedia lectures or HTML web pages and improves the quality of eLearning content retrieval. Then, Sandler et al [85] examined the potential of two ranking models (paragraph recovery model, dependency-based re-ranking) with varying purposes to improve the performance of learning-focused search engines. In their research, Syed and Collins Thompson [86], [87] proposed to optimize learning outcomes by selecting a set of documents while taking into account the keyword density and domain knowledge of the learner. Their theoretical framework provides a solid basis for the further study of learning-oriented recovery techniques. Lu and Hsiao [88] studied the information-seeking behavior of users in programming language forums. They also designed a personalized information retrieval assistant that promotes learning through the modeling of user behavior and query refinement, thus showing significantly improved learning efficiency. Karanam et al [89] present a model for predicting clicks on search results incorporating the user's individual level of knowledge in the corresponding domain. The authors then discuss and compare knowledge acquisition strategies suited for the current state of knowledge, showing significant gains in knowledge acquisition when using skill-based strategies. Pera et al [20] present an improved research environment, YouUnderstood.Me (YUM), which aims at supporting the learning of children from Kinder-

garten to Grade 9 by retrieving documents that do not only meet the information needs of children, but also their reading skills.

Remarks on the improved works on Search–As- Learning. We realize that current work for the improvement of search engines in a Search–As-Learning context proposes a reclassification (Sandler et al [85]), a refinement of the query (Collins Thompson [86], [87], Lu and Hsiao [88]) or an environment (Pera et al [20], Qiu et al [83], Bosnić et al [84]). Nevertheless, shortcomings are observed because the proposals do not highlight a filter to eliminate content for non-educational purposes to avoid cognitive overload. Also, we do not notice a reorganization of the results which takes into account the level of Bloom's taxonomy perceptible from the user's request. Also, with regard to the result extract, it was mentioned previously in Section 2.1.3 that it would be desirable to have on the search result page, extracts of text focused on the queries with a satisfactory length. At present, no work exploits these aforementioned mentioned components.

3 Discussion

Throughout this paper, we propose a state of the art around search engines in a learning context. Thus, we start with a presentation of some basic concepts on Information Retrieval (IR) because, it is largely the advances in IR that lead us to search engines today. Then, we highlight some aspects of the principle of search engines selection and ranking. Next, we present some relevant works on the display side of search engine results by examining: SERP and their presentation, generation of the snippet text, Number of results per page and snippet lengths.

All of these aspects around SERP lead us to understand that today's search engines are designed for all types of searches tasks. Indeed, the emphasis is not particularly on optimization in learning contexts. However, search for learning being informational search task, we note according to these previous researches that:

- snippets must be query oriented: so, remove irrelevant snippets (Ads, Events, etc.);
- snippets must have fairly informative content (6-7 lines);
- snippets should not be more than 10 results per page.

After doing this study around SERP, we turn this review around Search-As-Learning whose questions are more concise in terms of learning oriented search task. This presentation of works around Search-As-Learning categorizes researches according to the following themes: search tasks related to learning, exploratory search oriented learning, expertise and learning during search tasks, evaluation of learning during search tasks, optimization of search results for education.

Following the presentation of works around Search-As-Learning we note in our aforementioned remarks certain aspects which we consider important to take into account in the optimization of search engines for learning cases. These aspects are:

- consideration of the level of the learner in accordance with the Bloom's taxonomy when classifying the results;

- elimination of irrelevant contents for learning.

In fact, in our future works, we intend to optimize the search results with regard to educational objectives by relying on the remarks that we made in this literature review.

4 Conclusion

In this work, we provide a state of the art on search engine in a learning context. We started by recalling some information retrieval (IR) bases, then, we illustrated the search engine selection and display result policy. In terms of user interface, we brought out some research around SERP. After this literature review on contexts surrounding SERP, we brought specifically a state of the art on learning during search tasks. We note that Search-As-Learning addresses questions around human learning during the search tasks. Concerning the selection and ranking of algorithms in search engines, particularly the Search-As-Learning optimization algorithms, we realize that no algorithm is interested in determining the learner knowledge level from the request and the elimination of non-educational content from search engines. In addition, works on SERP agree that exploratory (informational) search should display fairly informative content in result snippets or summaries. Based on the state of the art, we can say that the advantages of informative snippets include the rapid assessment of the relevance of an extract on the results page, the time to time saving during the appropriation of the content, the ability for the user to better reformulate his requests during the search process, and the improvement of the readability of the extracts text. In our next research, our goal will be to provide a layer above the search engines that responds to two levels: algorithmic and interface levels. At the algorithmic level, we identify the level of the user from the request, and then eliminate the extracts of non-educational content (advertising, Tweets, Addresses, Maps, etc.) during the search to keep only snippets for educational purposes which will be classified according to the level identified in user request. At the interface level, we propose a result snippet model for educational content.

5 References

[1] Purcell, K., Brenner, J., & Raine, L. (2012). Search Engine Use 2012. Pew Internet & American Life Project. Van Eimeren, B., & Frees, B. (2012).

[2] Baeza-Yates RA, Ribeiro-Neto B. Modern Information Retrieval, 2nd Edition. Boston: Addison-Wesley Longman Publishing Co., Inc; 2011.

[3] Leif Azzopardi, Paul Thomas, Nick Craswell, Measuring the Utility of Search Engine Result Pages: An Information Foraging Based Measure, The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, July 08-12, 2018, Ann Arbor, MI, USA. https://doi.org/10.1145/3209978.3210027

[4] W. Kules, M. Wilson, M. Schraefel, and B. Shneiderman. From keyword search to exploration: How result visualization aids discovery on the web. Technical Report

1516920080208, School of Electronics and Computer Science, University of Southampton, 2008.
[5] Rose, D.E. and Levinson, D. (2004). Understanding user goals in Web search. WWW, 13–19.
[6] Edward Cutrell, Zhiwei Guan. What Are You Looking For? An Eye-tracking Study of Information Usage in Web Search. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'07). ACM Press, New York, NY, USA, 407-416; 2007. https://doi.org/10.1145/1240624.1240690
[7] Clarke, Ch., Agichtein, E., S. Dumais, White, R. W.: The Influence of Caption Features on Clickthrough Patterns in Web Search. In: SIGIR2007.
[8] D. Maxwell, L. Azzopardi, and Y. Moshfeghi. 2017. A Study of Snippet Length and Informativeness: Behaviour, Performance and User Experience. In Proc. of SIGIR. 135–144. https://doi.org/10.1145/3077136.3080824
[9] Marcos, M.-C., Gavin, F., & Arapakis, I. (2015). Effect of snippets on user experience in web search. In Proceedings of the XVI International Conference on HCI (pp. 47:1–47:8). New York: ACM. https://doi.org/10.1145/2829875.2829916
[10] Zabolotniaia M, Cheng Z, Dorozhkin E, Lyzhin A. Use of the LMS Moodle for an Effective Implementation of an Innovative Policy in Higher Educational Institutions. International Journal of Emerging Technologies in Learning (iJET), Kassel, Germany: International Journal of Emerging Technology in Learning. 2020; 15. https://doi.org/10.3991/ijet.v15i13.14945
[11] Faiz Bin Zafrah, Khalid Mohiuddin, Ahmed A. Ahmed, Abd Almajeed Alajab, Osman A.Nasr. Asynchronous learning techniques during the COVID 19 Pandemic. International Research Journal of Multidisciplinary Technovation. Volume 3, Issue 2, Year 2021. https://doi.org/10.34256/irjmt2127
[12] Hansen, P., & Rieh, S. Y. (2016). Recent advances on searching as learning: An introduction to the special issue. Journal of Information Science, 42(1), 3–6 https://doi.org/10.1177/0165551515614473
[13] Ujwal Gadiraju, Ran Yu, Stefan Dietze, and Peter Holtz. 2018. Analyzing Knowledge Gain of Users in Informational Search Sessions on the Web. In Proceedings of the 2018 Conference on Human Information Interaction&Retrieval (CHIIR '18). ACM, New York, NY, USA, 2–11. https://doi.org/10.1145/3176349.3176381
[14] Souvick Ghosh, Manasa Rath, and Chirag Shah. 2018. Searching As Learning: Exploring Search Behavior and Learning Outcomes in Learning-Related Tasks. In Proceedings of the 2018 Conference on Human Information Interaction&Retrieval (CHIIR '18). ACM, New York, NY, USA, 22–31. https://doi.org/10.1145/3176349.3176386
[15] Ran Yu, Ujwal Gadiraju, and Stefan Dietze. 2018. Detecting, Understanding and Supporting Everyday Learning in Web Search. arXiv:1806.11046 [cs] (June 2018). arXiv:cs/1806.11046
[16] Ran Yu, Ujwal Gadiraju, Peter Holtz, Markus Rokicki, Philipp Kemkes, and Stefan Dietze. 2018. Predicting User Knowledge Gain in Informational Search Sessions. In The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR '18). ACM, New York, NY, USA, 75–84. https://doi.org/10.1145/3209978.3210064
[17] N Bhattacharya, J Gwizdka. Measuring Learning During Search: Differences in Interactions, Eye-Gaze, and Semantic Similarity to Expert Knowledge. Paper presented at the Proceedings of the 2019 Conference on Conference Human Information Interaction and Retrieval, New York, NY, USA. 2019. https://doi.org/10.1145/3295750.3298926

[18] Georg Buscher, Susan T. Dumais, Edward Cutrell, The good, the bad, and the random: an eye-tracking study of ad quality in web search, Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval, July 19-23, 2010, Geneva, Switzerland. https://doi.org/10.1145/1835449.1835459
[19] Eickhoff, C., Gwizdka, J., Hauff, C. and He, J. (2017), "Introduction to the special issue on search as learning", Information Retrieval Journal, Vol. 20 No. 5, pp. 399-402. https://doi.org/10.1145/3209978.3210064
[20] Azpiazu, I. M., Dragovic, N., Pera, M. S., & Fails, J. A. (2017). Online searching and learning: YUM and other search tools for children and teachers. Information Retrieval. https://doi.org/10.1007/s10791-017-9310-1
[21] Sylvain Sagot. Contribution à la conception et à la maîtrise du processus de référencement Web. Autre. Université de Technologie de Belfort-Montbeliard, 2016. Français. ffNNT: 2016BELF0310ff. fftel-01492889f
[22] O. Andrieu, L. Bourelly, et Grifil, Réussir son référencement web: stratégie et techniques SEO. Eyrolles, 2016.
[23] M. A. Hingoro and H. Nawaz, "A Comparative Analysis of Search Engine Ranking Algorithms," International Journal of Advanced Trends in Computer Science and Engineering, vol. 10, no. 2, pp. 1247–1252, Apr. 2021. https://doi.org/10.30534/ijatcse/2021/1081022021
[24] J.-P. Lardy, « Méthodes de tri des résultats des moteurs de recherche », La lettre de l'URFIST, 2000.
[25] K. Mohan and J. Kurmi, "A Technique to Improved Page Rank Algorithm in perspective to Optimized Normalization Technique," International Journal, vol. 8, no. 3, 2017.
[26] H. Chen and S. Dumais. Bringing order to the web: Automatically categorizing search results. In Proc. 18th ACM CHI, pages 145–152, 2000. https://doi.org/10.1145/332040.332418
[27] S. Dumais, E. Cutrell, and H. Chen. Optimizing search by showing results in context. In Proc. 19th ACM CHI, pages 277–284, 2001.
[28] F. Chieriche i, R. Kumar, and P. Raghavan. Optimizing two-dimensional search results presentation. In Proc. 4th ACM WSDM, pages 257–266, 2011.
[29] Y. Kammerer and P. Gerjets. How the interface design influences users' spontaneous trustworthiness evaluations of web search results: comparing a list and a grid interface. In Proc. of the Symp. on Eye-Tracking Research & Applications, pages 299–306, 2010. https://doi.org/10.1145/1743666.1743736
[30] M. L. Resnick, C. Maldonado, J. M. Santos, and R. Lergier. Modeling on-line search behavior using alternative output structures. In Proc. Human Factors and Ergonomics Soc. Annual Meeting, volume 45, pages 1166–1170, 2001. https://doi.org/10.1177/154193120104501503
[31] N. Craswell, O. Zoeter, M. Taylor, and B. Ramsey. An experimental comparison of click position-bias models. In Proc. 1st ACM WSDM, pages 87–94, 2008. https://doi.org/10.1145/1341531.1341545
[32] T. Joachims, L. Granka, B. Pan, H. Hembrooke, and G. Gay. Accurately interpreting clickthrough data as implicit feedback. In Proc. 28th ACM SIGIR, pages 154–161, 2005. https://doi.org/10.1145/1076034.1076063
[33] H. Joho and J. M. Jose. A comparative study of the effectiveness of search result presentation on the web. In Proc. 28th ECIR, pages 302–313, 2006. https://doi.org/10.1007/11735106_27
[34] H. Ali, F. Scholer, J. A. Thom, and M. Wu. User interaction with novel web search interfaces. In Proc. 21st OZCHI, pages 301–304.

[35] J. He, P. Duboue, and J.-Y. Nie. Bridging the gap between intrinsic and perceived relevance in snippet generation. In Proc. of COLING 2012, pages 1129–1146, 2012.
[36] T. Landauer, D. Egan, J. Remde, M. Lesk, C. Lochbaum, and D. Ketchum. Enhancing the usability of text through computer delivery and formative evaluation: the superbook project. Hypertext: A psychological perspective, pages 71–136, 1993.
[37] L. Leal-Bando, F. Scholer, and A. Turpin. Query-biased summary generation assisted by query expansion. J. Assoc. for Info. Sci. and Tech., 66(5):961–979, 2015. https://doi.org/10.1002/asi.23222
[38] J. Pedersen, D. Cutting, J. Tukey, et al. Snippet search: A single phrase approach to text access. In Proc. 1991 Joint Statistical Meetings, 1991.
[39] A. Tombros and M. Sanderson. Advantages of query biased summaries in information retrieval. In Proc. 21st ACM SIGIR, pages 2–10, 1998. https://doi.org/10.1145/290941.290947
[40] R. W. White, J. M. Jose, and I. Ruthven. A task-oriented study on the influencing effects of query-biased summarisation in web searching. Info. Processing & Management, 39(5):707–733, 2003. https://doi.org/10.1016/S0306-4573(02)00033-X
[41] J. Kupiec, J. Pedersen, and F. Chen. A trainable document summarizer. In Proc. 18th ACM SIGIR, pages 68–73, 1995. https://doi.org/10.1145/215206.215333
[42] N. V. Spirin, A. S. Kotov, K. G. Karahalios, V. Mladenov, and P. A. Izhutov. A comparative study of query-biased and non-redundant snippets for structured search on mobile devices. In Proc. 25th ACM CIKM, pages 2389–2394, 2016. https://doi.org/10.1145/2983323.2983699
[43] D. E. Rose, D. Orr, and R. G. P. Kantamneni. Summary attributes and perceived search quality. In Proc. 16th WWW, pages 1201–1202, 2007. https://doi.org/10.1145/1242572.1242765
[44] T. Kanungo and D. Orr. Predicting the readability of short web summaries. In Proc. 2nd ACM WSDM, pages 202–211, 2009. https://doi.org/10.1145/1498759.1498827
[45] H. Bast and M. Celikik. Efficient index-based snippet generation. ACM Trans. Inf. Syst., 32(2):6:1–6:24, Apr. 2014. https://doi.org/10.1145/2590972
[46] A. Turpin, Y. Tsegay, D. Hawking, and H. E. Williams. Fast generation of result snippets in web search. In Proc. 30th ACM SIGIR, pages 127–134, 2007. https://doi.org/10.1145/1277741.1277766
[47] Q. Li and Y. P. Chen. Personalized text snippet extraction using statistical language models. Pattern Recogn. 43(1):378–386, Jan. 2010. https://doi.org/10.1016/j.patcog.2009.06.003
[48] M. Ageev, D. Lagun, and E. Agichtein. Improving search result summaries by using searcher behavior data. In Proc. 35th ACM SIGIR, pages 13–22, 2013. https://doi.org/10.1145/2484028.2484093
[49] D. Savenkov, P. Braslavski, and M. Lebedev. Search snippet evaluation at yandex: lessons learned and future directions. Multilingual & Multimodal Information Access Evaluation, pages 14–25, 2011. https://doi.org/10.1007/978-3-642-23708-9_4
[50] K. M. Svore, J. Teevan, S. T. Dumais, and A. Kulkarni. Creating temporally dynamic web search snippets. In Proc. 35th ACM SIGIR, pages 1045–1046, 2012. https://doi.org/10.1145/2348283.2348461
[51] J. Kim, P. Thomas, R. Sankaranarayana, and T. Gedeon. Comparing scanning behaviour in web search on small and large screens. In Proc. 17th ADCS, pages 25–30, 2012. https://doi.org/10.1145/2484028.2484093

[52] J. Kim, P. Thomas, R. Sankaranarayana, T. Gedeon, and H.-J. Yoon. Pagination versus scrolling in mobile web search. In Proc. 25th ACM CIKM, pages 751–760, 2016. https://doi.org/10.1145/2484028.2484093
[53] J. Kim, P. Thomas, R. Sankaranarayana, T. Gedeon, and H.-J. Yoon. Eye-tracking analysis of user behavior and performance in web search on large and small screens. J. of the Assoc. for Information Science and Technology, 2014. https://doi.org/10.1002/asi.23187
[54] G. Linden. Marissa mayer at web 2.0, November 2006. http://glinden.blogspot.com/2006/11/marissa-mayer-at-web-20.html
[55] A. Oulasvirta, J. Hukkinen, and B. Schwartz. When more is less: The paradox of choice in search engine use. In Proc. 32nd ACM SIGIR, pages 516–523, 2009. https://doi.org/10.1145/1571941.1572030
[56] B. Schwartz. The Paradox of Choice: Why More Is Less. Harper Perennial, 2005.
[57] D. Kelly and L. Azzopardi. How many results per page?: A study of serp size, search behavior and user experience. In Proc. 38th ACM SIGIR, pages 183–192, 2015. https://doi.org/10.1145/1571941.1572030
[58] M. Hearst. Search user interfaces. Cambridge University Press, 2009.
[59] T. Paek, S. Dumais, and R. Logan. Wavelens: A new view onto internet search results. In Proc. 22nd ACM CHI, pages 727–734, 2004. https://doi.org/10.1145/985692.985784
[60] M. Kaisser, M. A. Hearst, and J. B. Lowe. Improving search results quality by customizing summary lengths. In Proc. 46th ACL, pages 701–709, 2008.
[61] J. Kim, P. Thomas, R. Sankaranarayana, T. Gedeon, and H.-J. Yoon. What snippet size is needed in mobile web search? In Proc. 2nd ACM CHIIR, pages 97–106, 2017. https://doi.org/10.1145/3020165.3020173
[62] Bailey, P., Chen, L., Grosenick, S., Jiang, L., Li, Y., Reinholdtsen, P., Salada, C., Wang, H., and Wong, S. 2012. User task understanding: a web search engine perspective. In NII Shonan Meeting on Whole-Session Evaluation of Interactive Information Retrieval Systems, Kanagawa, Japan. (Oct. 2012).
[63] Eickhoff, C., Teevan, J., White, R., and Dumais, S. 2014. Lessons from the journey: a query log analysis of within-session learning. In Proc. of WSDM 2014. ACM, New York, NY, 223-232. https://doi.org/10.1145/2556195.2556217
[64] Raman, K., Bennett, P. N., and Collins-Thompson, K. 2014. Understanding intrinsic diversity in Web search: Improving whole.
[65] Jansen, B. J., Booth, D., and Smith, B. 2009. Using the taxonomy of cognitive learning to model online searching. Information Processing & Management. 45, 6 (Nov. 2009), 643-663. https://doi.org/10.1016/j.ipm.2009.05.004
[66] Lorin W. Anderson, David R. Krathwohl, and Benjamin Samuel Bloom. 2001. A Taxonomy for Learning, Teaching, and Assessing.
[67] Kim, J. Y., Collins-Thompson, K., Bennett, P. N., and Dumais, S. T. 2012. Characterizing web content, user interests, and search behavior by reading level and topic. In Proc. of WSDM 2012. ACM, New York, NY, 213-222. https://doi.org/10.1145/2124295.2124323
[68] Marchionini, G. 2006. Exploratory search: from finding to understanding. Comm. of the ACM. 49, 4 (April 2006), 41-46. https://doi.org/10.1145/1121949.1121979
[69] White, R. W., and Roth, R. A. 2009. Exploratory search: Beyond the query-response paradigm. Synthesis Lectures on Information Concepts, Retrieval, and Services. 1, 1 (2009), 1-98. https://doi.org/10.2200/s00174ed1v01y200901icr003
[70] Wildemuth, B. M. 2004. The effects of domain knowledge on search tactic formulation. J. of the American Society for Information Science and Technology. 55, 3 (Feb. 2004), 246-258. https://doi.org/10.1002/asi.10367

[71] White, Ryen W., Susan T. Dumais, and Jaime Teevan (2009), "Characterizing the Influence of Domain Expertise on Web Search Behavior," paper presented at ACM International Con ference on Web Search and Data Mining, Barcelona, Spain (February 9-12). https://doi.org/10.1145/1498759.1498819
[72] Zhang X., Cole M., and Belkin N. 2011. Predicting users' domain knowledge from search behaviors. In Proceedings of SIGIR '11. ACM, New York, NY, 1225-1226. https://doi.org/10.1145/2009916.2010131
[73] P. Vakkari. Searching as learning: A systematization based on literature. Journal of Information Science, 42(1):7–18, 2016.
[74] Vakkari, P., Pennanen, M. and Serola, S. 2003. Changes of search terms and tactics while writing a research proposal. Information Processing & Management. 39, 3 (May 2003), 445-463. https://doi.org/10.1016/s0306-4573(02)00031-6
[75] Vakkari P., and Huuskonen S. 2012. Search effort degrades search output but improves task outcome. J. American Society for Information Science and Technology. 63, 4 (April 2012), 657-670. https://doi.org/10.1002/asi.21683
[76] Hersh, W. R., Elliot, D. L., Hickam, D. H., Wolf, S. L. and Molnar, A. 1995. Towards new measures of information retrieval evaluation. In Proc. of SIGIR 1995. ACM, 164-170. https://doi.org/10.1145/215206.215355
[77] Kammerer, Y., Nairn, R., Pirolli, P., and Chi, E. H. 2009. Signpost from the masses: Learning effects in an exploratory social tag search browser. In Proceedings of SIGCHI 2009. ACM, 625–634. https://doi.org/10.1145/1518701.1518797
[78] Wilson, M. J., and Wilson, M. L. 2013. A comparison of techniques for measuring sensemaking and learning within participant-generated summaries. J. of the American Society for Information Science and Technology. 64, 2 (Feb. 2013), 291-306. https://doi.org/10.1002/asi.22758
[79] Huda, M., Maseleno, A., Atmotiyoso, P., Siregar, M., Ahmad, R., Jasmi, K., et al. (2018). Big data emerging technology: insights into innovative environment for online learning resources. International Journal of Emerging Technologies in Learning (iJET),13(1), 23–36. https://doi.org/10.3991/ijet.v13i01.6990
[80] Y. H. Woon, "Students' Perception about Learning using MOOC", International Journal of Emerging Technologies in Learning (iJET), vol. 14, no. 18, pp. 203-208, Sep. 2019. https://doi.org/10.3991/ijet.v14i18.10802
[81] Piyush Arora. 2015. Promoting User Engagement and Learning in Amorphous Search Tasks. In Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM, 1051–1051. https://doi.org/10.1145/2766462.2767848
[82] Kodama, C., St. Jean, B., Subramaniam, M., & Taylor, N. G. (2017). There's a creepy guy on the other end at Google!: Engaging middle school students in a drawing activity to elicit their mental models of Google. Information Retrieval. https://doi.org/10.1007/s10791-017-9306-x
[83] S Qiu, A Bozzon, U Gadiraju. Conversational Interfaces for Search As Learning - CIKM (Workshops), 2020.
[84] Bosnić, I.; Žagar, Mario; Žagar, Martin. In Search of a Smile: SMIL Indexing System for Multimedia Learning. International Journal of Emerging Technologies in Learning. 2007, Vol. 2 Issue 3, p5-11.
[85] M. Sandler, "Organizing search results in a topic hierarchy," 2012.
[86] Rohail Syed and Kevyn Collins-Thompson. 2017. Retrieval algorithms optimized for human learning. In Proceedings of the 40th International ACM SIGIR Conference on Re-

search and Development in Information Retrieval. ACM, 555–564. https://doi.org/10.1145/3077136.3080835

[87] Syed, R., Collins-Thompson, K. (2016). Optimizing search results for educational goals: Incorporating keyword density as a retrieval objective. http://ceur-ws.org/Vol-1647/SAL2016_paper_21.pdf

[88] Lu, Y. & Hsiao, I. H. (2017). Personalized information seeking assistant (PiSA): From programming information seeking to learning. Information Retrieval. https://doi.org/10.1007/s10791-017-9305-y

[89] Karanam, S., Jorge-Botana, G., Olmos, R., & van Oostendorp, H. (2017). The role of domain knowledge in cognitive modeling of information search. Information Retrieval. https://doi.org/10.1007/s10791-017-9308-8

6 Authors

Jaurès S. H. Kameni is a post graduate Computer Science Engineer and PhD student at National Advanced School of Engineering (ENSP) of the University of Yaoundé 1, PO Box 8390 Yaoundé-Cameroon. He is also Senior Developer for a Company in Cameroon. His research areas are eLearning and Machine Learning.

Bernabé Batchakui PhD. is Doctor /Engineer in computer Engineering and Lecturer at the National Advanced School of Engineering (ENSP) of the University of Yaoundé 1, PO Box 8390 Yaoundé-Cameroon. He is interested in artificial intelligence applied to education, with a strong emphasis on knowledge engineering for solving problems in developing countries. In addition, he is responsible of the team in charge of training teachers around MOOCs at the ENSPY. (Email: bbatchakui@gmail.com).

Roger Nkambou is full Professor, Department of Computer Science Faculty of Science University of Quebec at Montreal. He is former Director of the Centre de Recherche en Intelligence Artificielle: CRIA (Email: nkambou.roger@uqam.ca).

Article submitted 2021-08-13. Resubmitted 2021-10-29. Final acceptance 2021-11-04. Final version published as submitted by the authors.