Web Hypermedia Resources Reuse and Integration for On-Demand M-Learning

Jawad Berri[1], Rachid Benlamri[2], Yacine Atif[3] and Hajar Khallouki[2]

[1]Dept. of Information Systems, King Saud University, Riyadh, Saudi Arabia,
[2]Dept. of Software Engineering, Lakehead University, Canada
[3]University of Skövde, Skövde, Sweden

Abstract
The development of systems that can generate automatically instructional material is a challenging goal for the e-learning community. These systems pave the way towards large scale e-learning deployment as they produce instruction on-demand for users requesting to learn about any topic, anywhere and anytime. However, realizing such systems is possible with the availability of vast repositories of web information in different formats that can be searched, reused and integrated into information-rich environments for interactive learning. This paradigm of learning relieves instructors from the tedious authoring task, making them focusing more on the design and quality of instruction. This paper presents a mobile learning system (Mole) that supports the generation of instructional material in M-Learning (Mobile Learning) contexts, by reusing and integrating heterogeneous hypermedia web resources. Mole uses open hypermedia repositories to build a Learning Web and to generate learning objects including various hypermedia resources that are adapted to the user context. Learning is delivered through a nice graphical user interface allowing the user to navigate conveniently while building their own learning path. A test case scenario illustrating Mole is presented along with a system evaluation which shows that in 90% of the cases Mole was able to generate learning objects that are related to the user query.

Key words:
On-demand learning, learning object, web hypermedia resource, resource integration, mobile learning system.

1. Introduction

Web 2.0 has created many opportunities for users who have mutated from information consumers –searching and browsing, to information producers –interacting and creating information. Users are becoming more active in dealing with web content; they are expecting more from this open space that has changed their way of life [1]. This change in the user's role from consumer to producer, has enticed research in many directions to investigate how existing information can be leveraged to create smart web applications that aggregate existing information to create and manage knowledge. In this context, Web 2.0 Learning is emerging as a new paradigm of modern education [2] [3]. The web has the potential to feed e-learning since the content provided to web users is presented in hypermedia form including text, data, picture, animation, audio, video, presentation and others [4]. This hypermedia information available on the web is increasing everyday, offering the biggest information repository in the world. Finding the relevant information is however a challenge task for millions of users who access the web daily using web search engines. Although this information is very rich in its content and form, it lacks a formal description that allows users to search more conveniently using intuitive techniques such as natural language. In fact, existing search engines are still not able to correlate a query with different types of multimedia information from different web repositories. Reusing and integrating this rich information to generate instructional material on-demand is a real challenge for the e-learning community.

Technologies that are associated with this change such as mobile smart phones and tablet computers, and the availability of wireless networks in urban spaces, are also creating new forms of interaction. These technologies add a new interaction dimension to hypermedia challenges, as they are enabling new services to satisfy nomadic user needs in terms of information, entertainment and social networking [5] [6]. The digital learning scene engendered by Web 2.0 and Information and Communication (ICT) technologies is thus becoming mobile, context-aware, collaborative, participatory, and non-linear.

This paper presents Mole, a mobile learning system that reuses available hypermedia web resources and integrates them to generate instructional material in response to user's queries [7] and their context. Users will be able to learn through a course structure generated on demand from the web allowing them to be exposed to a succession of organized learning sessions. Web resources are retrieved and semantically annotated in order to be utilized in Mole, allowing users to learn within a given context. Hypermedia web resources combining multimedia documents and hyperlinks are particularly interesting as they promote the development of interactive and non-linear material for learning. Mole architecture is designed as a set of modules that i) gather and organize multimedia information from the

Manuscript received January 5, 2021
Manuscript revised January 20, 2021
https://doi.org/10.22937/IJCSNS.2021.21.1.17

web; ii) correlate and package the content into lightweight Learning Objects (LOs) according to pre-set layouts; iii) integrate LOs into a learning web (LW) – a learning course structure; and finally, iv) deliver the resulting interlinked LOs to the user allowing him/her to navigate through the learning web according to their needs and constraints. Mole generates learning sessions in response to mobile user queries. It uses open multimedia repositories to build a learning web and to generate learning objects including various hypermedia resources. Learning is delivered through a usable graphical user interface allowing the user to navigate conveniently, thus building a personal learning path.

2. Related Work

Lots of initiatives and works have considered the web as a source of learning. Instructors are keen to use tools, web services and web 2.0 technologies such as blogs, wikis, and other social networking applications to support the creation of ad-hoc learning communities [8]. These experiences led to learning environments where learners collaborate and participate actively in the learning activities [9]; they become motivated and more involved. For instance, the social network Facebook is frequently used in classroom as it offers a familiar virtual environment. Students can create communities in Facebook to exchange documents, educational resources and ideas, besides interacting with each other. On the other hand, the classroom teacher can regulate the learning sessions in a fashionable way. A special issue edited by Sampson and al. on current advances in learning technologies [10] has included various contributions that are representative of technologies used currently in technology-enhanced learning. The diversity of technologies presented in this special issue such as cloud computing for education, advanced assessment technologies, digital gaming technology, intelligent learning systems, and other technologies shows that e-learning can encompass new computing technologies for learners' benefit. This shows also that the e-learning community and specifically instructors are keen to exploit the new technology innovations in order to design intuitive instructional material.

Dunlap and Lowenthal presented in [11] how educators can use Web 2.0 technologies associated with blogging, social networking, document co-creation, and resource sharing to help students develop the skills needed to be effective lifelong learners. Web 2.0 technologies are used to help students develop fundamental skills such as autonomy, responsibility, and intentionality. They promote engaging students into communities of practice and encourage them to participate in discourse and collaboration. The work in [12] presents an online interactive platform APPRAISALWeb that analyzes, classifies and evaluates web resources that are used in language learning from a pedagogical point of view. This work is interesting as it focuses on the pedagogical evaluation of resources as a prerequisite for their use in e-learning. Brusilovsky et al. [13] propose an architecture for e-learning systems that is based on reusing adaptive content services. The authors advocate that the methods and tools developed for courseware re-use systems and adaptive Web-based educational systems can contribute to creating better Web-enhanced instruction. The Semantic Web community is also developing frameworks to make the meaning of learning material explicit and more accessible to deal with. The idea is to use ontology technologies to assist developers and educators organizing, personalizing, and publishing learning content. Pahl and Holohan in [14] discuss the potential of this technology for the development of learning content by using annotations and metadata. Learning content is described through metadata to allow the publication and discovery of these resources. Annotations can help to link these resources to underlying knowledge, thus making the knowledge explicit. In an attempt to develop a system to compose and package learning objects, Atif et al. in [15] [16] have proposed an interoperable implementation infrastructure using standard specifications for distributed learning system. The infrastructure relies on an extension of the Learning Object Metadata (LOM) specification for learning objects to integrate learning-adaptivity features into the learning object construct. The approach proposes a procedural methodology for LO construction as well as a hierarchy navigation of learning units that involves LOs.

Most systems that used the web as a support for learning consider in general the web technologies as a medium to enrich interaction and participation of learners. Alternatively, other systems used web resources, either embedded manually or referred by hyperlinks, as a support for their courses. In this paper we show that web resources are searched, reused and integrated automatically in learning objects to formulate e-learning courses that are generated on-demand to users.

3. Learning from Web Information

From the early days of Internet, web search engines have been developed to deliver information. Search engines are based on information retrieval principles, where mainly terms are used to index web documents. Term frequency measures are an indication of the document relevance when matched with user's query containing the same term. When a user accesses web pages retrieved by a search engine, he/she implicitly develops knowledge as he/she goes throughout the returned results. In fact, learning occurs when users incorporate retrieved information into their cognitive experience. Although learning occurs sporadically in this situation, the presented information suffers from instructional-shortcomings preventing a full interactive learning experience, such as: i) results are in the form of unrelated pieces of information retrieved from many web sites leaving the responsibility to the user to filter-out inappropriate material; ii) users do not have the opportunity to get a deeper knowledge relevant topics of interest, in which case they may need to undergo several search cycles to reformulate and refine further their queries; iii) different search engines retrieve for the same search query a variety of media results, and the user needs to further specify the media type from the expected results (web, images, maps, videos, news, presentations or documents); and iv) web search engines deliver a single session to users with no recorded history; they do not keep a record of previous searches nor they propose a management of sessions to allow users to visit their search results through a sequence of episodes.

Mole, the system, proposed in this paper, suggests a refined approach to present instructional information. Instead of presenting information as a list of ranked web pages as web search engines do, Mole proposes a learning environment where users navigate through a prearranged presentation of hypermedia resources related to the user's topic of interest. Hence, Mole relies on a redeployment of web-based hypermedia resources which are reorganized semantically to allow users to learn and build knowledge in a disciplined and guided approach. This approach of dealing with information to generate learning material is characterized by five main features:

1. Reusing existing web hypermedia resources: Mole does not create learning resources but reuses existing web resources that fit the user query.
2. Integrating diverse forms of hypermedia information: this integration promotes user's interaction as it involves different cognitive levels of learning.
3. Allowing learners to construct their personal learning path: the learning path represents the particular learner experience that he/she follows among the proposed learning structure. This way of structuring personalized learning experiences promotes adaptive and non-linear instruction to accommodate a variety of learning styles and background-levels.
4. Adapting the content to suit the learner context: as mobiles perceive contexts, the learning material proposed to learners could be adapted to the learner current environment and/or activity.
5. Managing learning through sessions: storing the progress of learners is very important to allow learners to control the amount of material they wish to be exposed to, given their pace and constraints, and also to be able to resume learning at any time.

Mole promotes e-learning on-demand where learners regulate the pace and depth of learning. This just-in-time paradigm of learning delivers just-enough-learning for effective retention [17].

4. Learning Object Generation from the Web

The authoring task for a learning system is time-consuming, requiring from the instructor to develop their own material and/or re-use educational resources that are open access on the web. Using the web to generate automatically learning objects is a very motivating idea as it is poised to relieve instructors from systematically authoring instructional material for a given courseware, which is a persistent burden. Instructors are instead expected to query the web and get back learning objects that they can integrate into their courseware. Nowadays, this has been possible with the availability of standard specifications such as [18] and [19], allowing sharing and re-use of learning objects from available open-web repositories. However, learning material generated in such a way lacks instructional design as well as depth in terms of scope and coverage. Huge efforts need to be deployed in order to author LOs for every topic that could be requested by learners and satisfy their learning needs and context.

In this paper, we propose an architecture that is able to generate automatically learning objects from the web. Learning objects are composed of multimedia resources retrieved from different web repositories, which are re-organized and then packaged into courseware structures. In order to build useful LOs, web multimedia resources are expected to be described semantically so that semantically-correlated resources can be combined to form consistent knowledge for the targeted learning experience.

4.1 Semantic correlation function for web multimedia resources

With the availability of Web 2.0 tools and the popularity of mobile devices, web users became more active participants in authoring multimedia resources, which can be made available online. For instance, videos in shared repositories such as YouTube have been authored by thousands of users. As a consequence, descriptions of these resources were not written homogeneously. In order to be able to use multimedia resources in a learning system, these resources must be described semantically in a formal and a consistent way [20]. The Semantic Web initiative has proposed formalisms such as the Resource Description Framework (RDF) for this purpose [21] [22]; though it assumes that authors describe their proposed resources using these semantic specifications. Annotating resources manually requires coordinated efforts, enormous manpower and means; it is simply not feasible. Another struggle is the format of these resources: the majority of multimedia resources are binary files that cannot be processed automatically to extract semantic features such as names of persons, type of activities, nature of events, etc. Natural language techniques can help annotate and semantically describe these resources. For an accurate semantic annotation, we propose to use a linguistic method that analyses the resource's description and related data. We propose a semantic correlation method that establishes semantic associations between multimedia resources based on semantic annotations and metadata information.

4.2 Contextual exploration for semantic annotation

Contextual Exploration (CE) is a method that can greatly contribute to solving the tricky problem of computing semantic correlations among multimedia web resources. CE has been developed to solve many problems related to natural language processing. It is a decision-based method that does not require huge linguistic resources and fine descriptions of linguistic units. Instead, it involves only the linguist grammatical and lexical know-how regarding decision-making, when solving a linguistic problem [23]. The principle of the method is similar to the behavior of a human who is reading a text to analyze it in view of taking a decision. CE scans first a linguistic context looking for linguistic markers that can trigger decisions. Markers can be any word occurrence, morpheme, lexeme, or lexical unit. Once a marker is found, the context is further analyzed to find contextual clues surrounding the marker to reinforce taking an unambiguous decision. The method relies on linguistic markers and clues that are organized into a database to handle texts. CE fits very well with unrestricted texts such as those written by web users to describe their authored multimedia resources and uploaded on the web. It can annotate semantically existing multimedia resources by analyzing the descriptions and also the associated comments added by users. For instance, if a user would like to know the date in which the famous Mona Lisa has been painted, he/she submits the following query: "*When was Mona Lisa painted*". It is clear that the user is looking for a "time information" as he/she has used the adverb "When", thus CE will annotate this query as "time information" request. When submitted to YouTube, it is not the first retrieved video which displays the information about time, but it is the second. In the author's description of the second video the following sentence is found: (…) *and is believed to have been painted between 1503 and 1506*. This sentence can easily be annotated by CE as "Year" as it contains a number between 0 and 3000. CE will match between the query and the sentence in the description since "Year" refers to years in the Gregorian calendar, which is an information about time. Accordingly, CE will select the second video as a reply to the user's query. In order to annotate multimedia resources, CE relies on linguistic knowledge which is gathered through a knowledge acquisition phase. Some language expression patterns that will be tagged as "time information" are presented in Table 1.

Linguistic Expression Patterns [24] are expressed in the EBNF (Extended Backus-Naur Form) notation where capitalized words are non-terminals. For instance, the non-terminal *ListTimePeriod* refers to the following list of words: *millennium, century, decade, year, month, week, day*.

CE annotates resources by processing their descriptions since resources (as binary files) such as videos, animations, audios and images cannot be handled semantically. The procedure correlating resources uses the semantic annotations to match semantically between resources and the user's query.

Table 1: Time linguistic patterns

Time Information	
Linguistic Expression's Pattern	Examples
[Numeral + Month + Year]	13 October 2018, September 2020
[Numeral + ListTimePeriod]	20^{th} century, 15 years, 4 weeks
[Year]	1998, 2020
[ListRefDays]	tomorrow, yesterday, morning, sunset

4.3 Matching algorithm

In order to match between resources and the user's query, CE uses a matching algorithm which matches linguistic patterns to sentences [25]. The algorithm is designed and implemented to be completely decoupled from the data so that to allow the system to be updated smoothly. Besides, the algorithm includes three pattern matching modes namely: the *Pure Sequential*, *Sequential* and *Random* matching modes to offer a high flexibility in matching natural language expressions and allow to lose or tighten the search.

The Pure Sequential search mode is a constrained search as words in the linguistic expressions are matched with text tokens with respect to their consecutive order. Also, no intermediate text tokens are allowed in between the text tokens matched for this search mode. The Sequential search mode requires that the matched words are in sequence but accepts the presence of other tokens inside the matched expression. The Random search mode necessitates the presence of all the words of the pattern in the sentence but has no restriction on their order.

In order to organize the search, the algorithm checks patterns according to their lengths (number of words in the pattern). Accordingly, the most restrictive patterns, the ones with highest length, are checked first. If no result is found, the algorithm considers the lower length patterns. This organization of patterns came up from experimentations which show that the larger the length of the pattern is, the accurate is the tagging. Lower length patterns generate sometimes noise that is irrelevant in the tagging process.

4.4 Learning context modeling

An ontological design is adopted in this study to model the user learning context. The proposed context ontology is composed of different classes as shown in Fig. 1. The device class collects the description of all user devices. For each device, it gives the user's hardware and software information and characteristics like screen resolution, CPU speed, memory capacity, battery level, OS name, the set of applications running in the used device, etc. The UserProfile class contains user related personal information, such as their spoken language, location, etc. The materialType class refers to the nature of the LOs (e.g., video, audio, etc.) selected for the learning process. Finally, the AdaptationsServices class describes different adaptive services, such as *Transcoding service* that enables the conversion of format (e.g., JPEG to PNG); *Transmoding service* that enables the conversion of types (e.g., text to sound); and *Transforming service* that allows the content change without changing the media type and format (e.g., text summarization, language translation, etc.)

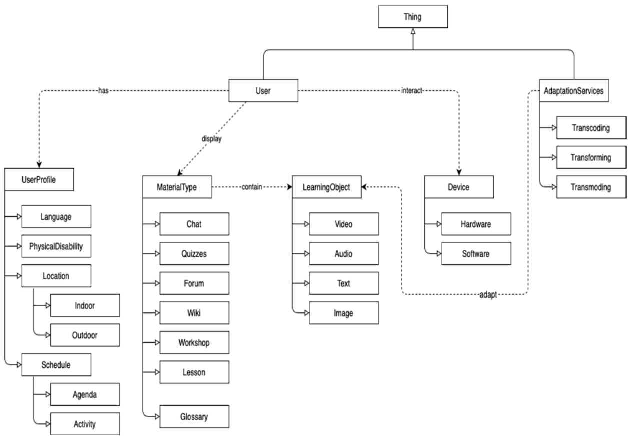

Fig. 1 Context modeling ontology.

To extract the data stored in an ontology, we use SPARQL language. The following SPARQL query shows how we retrieve the user location using the proposed context ontology. The various user context adaptation services proposed in this study are expressed by reasoning rules, which are encoded in the Semantic Web Rules Language (SWRL). The use of SWRL rules improves the expressiveness of the proposed context ontological model as illustrated in Figure 2.

SPARQL Query

SELECT ? location
WHERE{
?User rdf:type ont:John
?User ont:locatedIn ?location
}

SWRL Rules

Rule 1:
[*user(?x, John) ^ display(?x, audio1) ^ participate(?x, meeting1) → adaptationService(Transmoding)*]

Rule 2:
[*user(?x, John) ^ speak(?x, english) ^ displays=(?x, text1) → adaptationService(Transforming)*]

Rule 3:
[*user(?x, John) ^ locatedIn(?x, car) ^ display(?x, text2) → adaptationService(Transmoding)*]

Fig. 2 Examples of SPARQL query and SWRL rules for context adaptation.

For instance, the first rule indicates that if the user is participating in a meeting and he/she would like to display an audio content, a transmoding service will be triggered.

Algorithm: *LO Adaptation*

Input:
owl: ontology //the knowledge representation
IR: Inference Rules
LO: LO //Multimedia document to be adapted
Output:
AMD: Adapted Multimedia Document
{
r=NULL;
r=proceedAdapt(owl, IR, MD); //r is the result of the adaptation reasoning
AMD=AdaptationService(r); //Proceed the adaptation result based on reasoning result
return AMD; //return adapted multimedia document
}

Fig. 3 LO adaptation algorithm.

The LO adaptation algorithm, described in figure 3, adapts the learning object to the user context, making use of the set of predefined context inference rules. The process of the adaptation is realized through the method *proceedAdapt()* that takes as parameters the ontology (owl), the inference rules and the learning object to be adapted. The result of the *proceedAdapt()* method will be exploited by the *AdaptationService()* function in order to perform the adaptation process. *AdaptationService()* method relies on data sensed from user devices. Semantic annotations of sensed data are embedded within the various sensing stations employed by the user, and represented by snippets. Snippets are plain JSON objects that are embedded directly within the user's sensing devices. Figure 4 shows the snippet for sensing feature related to GPS sensor. The "sensor" identifies the GPS sensor, while the "observes" indicates the observed property. The snippet for sensing data is composed of an observed property and an observed value.

The "sensorOutput" identifies the observed property, while the "observedData" indicates the observed value which contains the longitude and latitude. Below, we introduce the semantic description generated by the ThingSpeak platform.

The latter is an IoT analytics platform service that allows a user to aggregate, visualize, and analyze live data streams about its context in the cloud. One can send data to ThingSpeak from various devices, create instant visualization of live data, and send alerts.

Once the user's context data is sensed, the semantic annotations become semantic descriptions encoded through the JSON-LD notation. According to the Snippet, a sensor of location is presented as shown in Figure 4, where the sensor value becomes the description @id; the identifier for this object is gps_shield. The observes value becomes the description @id; the identifier for it is location.

Location sensor Snippet

```
{
"sensor": "gps_shield",
"observes": "location"
}
```

Location observation Snippet

```
{
"sensorOutput": "location"
"observedData": {
"latitude": "33.030908",
"longitude": "-7.616505"}
}
```

Location sensor description

```
{
"@id": "gps_shield",
"@type": "CaMaOntology: sensor",
"observes": {
"@id": "location",
"@type": "CaMaOntology: location"}
}
```

Location observation description

```
{
"@id": "location",
"@type": "CaMaOntology:location",
"observedData": {
"latitude": {
"@id": "CaMaOntology:latitude",
"@type": "xsd:double",
"@value": "33.030908"}
"longitude": {
"@id": "CaMaOntology:longitude",
"@type": "xsd:double",
"@value": "-7.616505"}
                    }
}
```

Fig. 4 Snippet and sensor descriptions.

Figure 4 shows the location observation description which provides the GPS coordinates; the latitude and the longitude that are identified by their @id, data type and their values. Finally, the data mapping process between the JSON-LD description that contains the sensed data, and the OWL description that describes the user context, is performed by *JSON-LD_to_OWL* algorithm as shown in Figure 5. The algorithm takes as inputs the JSON-LD and the OWL descriptions. It starts by getting data from JSON-LD, then checks if JSON-LD description contains the required sensed values. If yes, then the owl description will be filled with the collected values

Algorithm: *JSONLD to OWL*

Input:
jld: JSON-LD file, owl: owl file
Output:
owl: final owl file

```
{
sensor=jld.getSensor();       //get sensed data from JSON-LD
lat=jld.getLatitude();
long=jld.getLongitude();

//testing if json-ld contains data collected from sensors
if lat != "" and long != "" and sensor != "" then
   owl.setSensor(sensor);     //fill sensed values into Owl
   owl.setLatitude(lat);
   owl.setLongitude(long);
end if

return owl;
}
```

Fig. 5 JSON-LD_to_OWL algorithm.

5. System Architecture

The system architecture is presented in Figure 6. The architecture depicts four modules that are necessary to find relevant information from the web in order to deliver learning resources to users. The system is based on the client-server model. On the client side, the user types a query which is interpreted as a request to learn about a topic. On the server side, the query is analyzed to allow searching for multimedia resources which are then packaged as learning objects and presented to the user. We describe below the main components of the proposed system.

5.1 Web 2.0 interface

Multimedia information available on the web needs to be accessed, processed and used within learning-objects generated for the user. Web 2.0 Interface is a module that is an interface to Web 2.0 technologies of interest in terms of multimedia content such as video and image repositories, wikis, blogs, presentations, maps, etc. Most of well-known multimedia repositories offer a convenient Web API

(Application Programming Interface) for web users to search and use multimedia resources publicly free of charge. APIs are in principle similar to web services offering a set of Hypertext Transfer Protocol (HTTP) request messages, however most of the Web 2.0 technologies have been moving away from Simple Object Access Protocol (SOAP) based services towards more complex querying command languages. As a result, APIs are not standardized since they are developed by different providers. Hence, in order to access multiple repositories, we need to implement a specific programming interface that invokes resources using an API for each service provider.

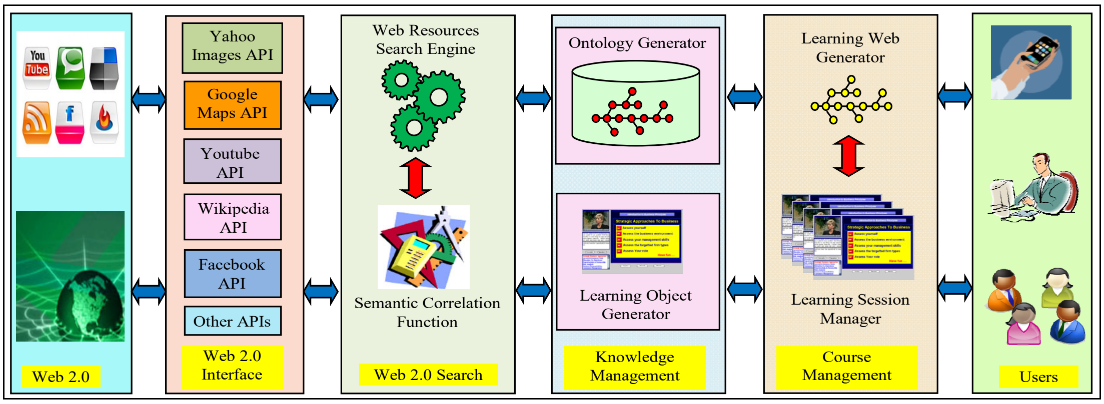

Fig. 6 System architecture.

5.2 Web 2.0 search

The search functionality is very important as it scouts the web to find suitable hypermedia resources that correspond to a user's query. The search module includes a web-resource search engine and a semantic correlation function, as discussed in previous sections. The web-resources search engine is an information retrieval system that focuses on hypermedia resources. An example of such systems are Federated Search Engines (FSE) such as Dogpile (dogpile.com) and Biznar (biznar.com). FSEs are information retrieval systems that work like a hub for many classical search engines. A query submitted to FSE is distributed to search engines, databases or other query engines participating in the federation and results are displayed through a single user interface. A major advantage of these search engines is their ability to focus the search on only a specific type of multimedia resources. The semantic correlation function computes semantic similarity between hypermedia resources. This function relies on natural language processing techniques that analyse texts associated with every web resource in order to compute the semantic similarity between the query and web resources.

5.3 Knowledge management

Most learning systems are based on ontologies in order to extract key learning concepts queried by the learner, as well as the sequence structure to organize the learning, mainly referred to as the learning web. Many ontology repositories are available on the web, such as the ontology repositories of W3C (http://www.w3.org/wiki/Ontology_repositories) and the open ontology repository initiative (http://ontolog.cim3.net/cgi-bin/wiki.pl?OpenOntologyRepository). These repositories are providing ready-made ontologies for many domains that are publicly available for use. These ontologies are developed using standard languages such as OWL and RDFS which allow them to be integrated within any platform. Numerous works are done to make this integration possible; Qu and Cheng [26] propose a practical search engine to facilitate the search of ontologies from the web. In order to generalize this to all ontology repositories in the web, Viljanen et al. [27] are proposing a unified API to access all ontology repositories that are available on the web, which greatly facilitate sharing and reuse of ontologies through a universal method. Unfortunately, ontologies are not available for every topic that could be requested by the user. In order to cope with this shortage, the proposed system uses a knowledge structure which replaces the ontology and which can be constructed from the web. The idea is to use the table of content of DBpedia which is a good knowledge structure that can be used to organize concepts of a given topic. The Course Generator Structure of Fig. 6 is the module that is responsible of generating the knowledge structure that is used by Mole to guide users in their navigation.

DBpedia (http://wiki.dbpedia.org/) is the knowledge base that structures information of Wikipedia, the free encyclopedia (www.wikipedia.org). DBpedia allows querying Wikipedia with sophisticated queries to extract structured information. Wikipedia is the largest open encyclopedia on the web developed using the Wiki principle, where users (volunteers) collaborate to write articles. It has become the largest and most popular general reference work on the Internet with 30 million articles and more than 5 million articles in English [28]. Wikipedia is an interesting source of knowledge for learners. We adopted Wikipedia as a main information source in our system for the following reasons:

- DBpedia provides a semantic indexation of Wikipedia contents
- It is a comprehensive repository of information which changes dynamically and grows everyday encompassing all new topics of interest for millions of users,
- Information included in Wikipedia is structured as it is written according to a template; every page includes a table of contents that are hyperlinks to the sections and subsections within the page,
- Wikipedia's pages are rich in content including descriptive text, images, citation references, hyperlinks to other pages, references to video and audio files,
- DBpedia provides an API to access the contents of pages for using them in web applications.

The table of contents of the Wikipedia's pages is an interesting document structure that we use to structure the learning content presented to the user. The table of contents is the physical structure of the page that organizes sections, sub-sections and paragraphs. In general, it is also a conceptual structure that organises the page topic and its related sub-topics. The table of contents is used to generate the learning web, which is the navigation structure for our mobile learners. Although the table of contents is used in Mole, it cannot replace the ontology which is an ideal conceptual structure for learning a topic in a formal way. However, we have opted for this solution to avoid constraining possible queries the learner might submit. Indeed, should we choose to use web services for retrieving the ontology corresponding to the user query's topic, we would restrict the user to only the topics for which an ontology is available.

Wikipedia is used also to retrieve the text describing the topic of the web page as the first paragraph is a definition of the topic. Also, the image related to the first paragraph is retrieved and added to the LO. For pages where there is no image, the system uses Yahoo images to retrieve an image corresponding to the topic.

The second element of the knowledge management component is the learning object generator. The LO generator selects resources gathered by the web resources search engine which have been ranked semantically in order to package them into a learning object. This module uses learning object templates that are filled with web resources to produce learning objects. LO templates are assigned to users according to their profile and the current context. However, this functionality is outside the scope of this paper and subject to future developments.

5.4 Course management

The course management module shown in Fig. 6. constructs the learning web which is used to organize learning objects graph and to monitor learning sessions. This module includes two main components: the learning web generator and the learning session manager.

The learning web generator produces a learning web from the retrieved DBPedia structure related to the concept queried by the user. The learning web is a tree-like structure where nodes are learning objects representing the concepts forming the main constituents of the DBPedia structure. The learning web allows learners to navigate through learning objects by providing commands to move forward (normal progression in the learning web), backward (review a LO in the LW) or to request additional learning objects (sub concept of the current concept) to view a LO that is more specific than the current LO and get more understanding about a given concept. In practice every learner constructs his own learning path that matches his personal navigation in LW.

Learners progress across LW through learning sessions that are managed by the learning session manager module. While the user queries the system, he/she will be exposed to the first LO in the LW and then, may move to other LOs. If the user quits, he/she is prompted to save his current session. The session keeps the navigation history of the learner that is saved as a file representing the traversed learning path, and which includes all the LOs that have been visited by the user. Learners can at any time retrieve their previously saved sessions to resume their learning experience.

6. Mobile Learning System – Case study

In this section we present Mole, a mobile learning system that constructs a learning web and provides learning sessions to mobile learners [29]. Mobile learning is a new form of learning delivered through mobile handsets. It is very attractive as the learner is reachable all the time. Also, it is a personalized form of learning as the mobile handset is personal and includes data about the learner which allows learning material to be tailored to fit each learner's context. It is this context awareness consideration about the user's environment (such as location and time), that is used to adapt learning material. On the other hand, learning delivered through mobile handsets suffers from user constraints and mobile limitations. As for the user who is on move, the time he affords for learning is restricted. In general, the type of learning suitable for a nomadic learner can be assimilated to a "first-aid" learning with short-time sessions. Limitations due to the device are mainly its screen size which can accommodate a very limited number of multimedia items. Other restrictions depend on the mobile handset model such as the embedded memory size and the operating system potentials as well as available applications.

We first discuss a case study scenario describing a learning session with special emphasis on user interactions. We then reveal some results of the system in terms of accuracy of the retrieved LOs from search engines.

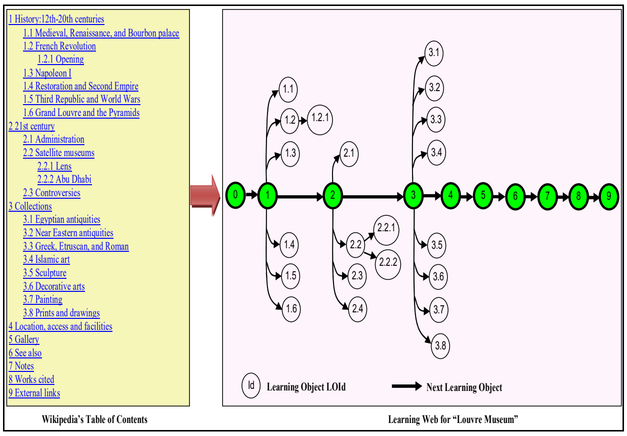

Fig. 7 Generating the learning web from Wikipedia table of contents.

In order to show the usage of Mole, we illustrate it through a case-study scenario. One of the typical situations where the user may solicit Mole is to learn about sightseeing while visiting a country as a tourist. Ayman is a fan of civilizations and does not miss an opportunity to visit museums all around the world. During a business trip to France, he decided to visit the Louvre Museum in Paris. Ayman had prior knowledge about the museum but he never visited it. Although he has a busy agenda, he could dedicate some time to make a visit to the museum. In order to get a background knowledge about the museum and to plan his visit, he requested a short learning session from the system to optimize his museum tour time. Through the welcome screen installed in his handheld device, he typed and then submitted the following query: "louvre museum". The system first forwards the query to Wikipedia where the page "louvre museum" is found. It retrieves the table of contents which is used as the learning web for delivering a learning session to Ayman (Fig. 7). Note that the identifiers inside the circles are DBpedia tags shown in Figure 7 corresponding to the section numbers of the table of contents in Wikipedia.

The circles in LW represent learning objects. The backbone of LW includes learning objects, which correspond to the main sections (numbered 1 to 9) of Wikipedia page. The first learning object presented to Ayman is LO0 that introduces the Louvre Museum (Fig. 8).

The text at the top left-side of LO0 corresponds to the first sentence extracted from the first paragraph of Wikipedia page; it is generally a definition of the page's topic. The user can click to have additional text about the current topic. The image besides the text is extracted from Wikipedia as for LO0. In case there is no image corresponding to the paragraph, Mole retrieves the first image from Yahoo images as for LO3 [30]. The video displayed in the LO is also the first video retrieved from YouTube as a response to the query. Yahoo images and YouTube services offer suitable APIs which allow Mole to retrieve and embed hypermedia resources in the generated LOs. "Main Topics" includes references to the learning objects corresponding to the main sections of Wikipedia page. "Sub Topics" are references to learning objects that correspond to the sub sections of the current section. In learning object LO3: Collections (Fig. 8), there are eight learning objects proposed to the user. The two arrows ("Next" and "Previous") at the bottom of the learning object allow the learner to move forward and backward to the learning objects at the same level of hierarchy in the LW.

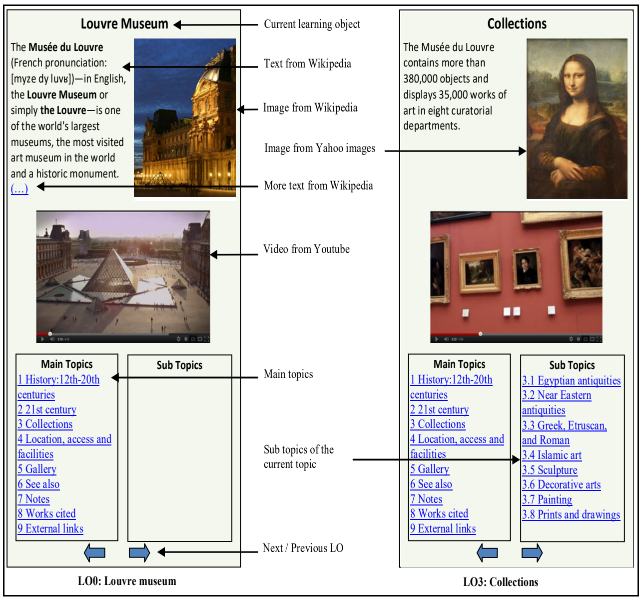

Fig. 8 Learning Object example.

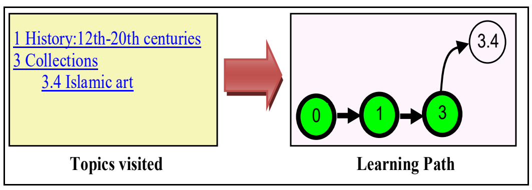

Fig. 9 Learning Object example.

The recorded navigation of Ayman shows that after LO0, he has clicked on "Next" to learning object LO1:

History: 12th-20th centuries, then he jumped to Learning Object LO3: Collections, and finally, he showed interest in Islamic art by moving to learning object LO3.4: Islamic art. The learning path generated by Ayman's navigation is shown in Fig. 9.

7. Evaluation

Mole has been tested with real-life user queries to evaluate its response quality and to explore potential improvements. Google Trending Searches[1] (GTS) service has been used to gather the set of queries GTS is an ideal choice for Mole as it provides real life user queries representing the most popular searches on Google which can be considered for Mole as typical queries for users who would like to learn about a topic of interest. GTS proposes a categorization of hot trend queries by country, language, time ranges and topics. We have used this service to extract queries between September 2019 and December 2020 for all topics and countries. We obtained a set of 325 queries out of which 263 were selected. This selection was based on two criteria: i) avoiding single word queries as phrases containing two and more words or acronyms are more interesting semantically and more challenging for testing Mole, and ii) having a mixture of topics to test Mole's responses with a diversity of user queries including people names, sport events, politics, art and cinema, health, food, etc.

7.1 Experiment setup

The experiment consisted in querying the three websites from which information is gathered for Mole (Wikipedia, YouTube and Yahoo images) and analysing the responses for every query. Evaluation of the results has been done by three evaluators[2]. First the evaluator accesses the full description of the hot trend in google corresponding to the query and gets an understanding of the subject related to the query. For each query in GTS, there is an image, a one line descriptive text and a hyperlink which refers generally to a newspaper article or a related website allowing the reader to have more information about the query's subject. The evaluator follows the hyperlink and reads the article, he is then exposed to the results of the three websites (Wikipedia, Youtube and Yahoo images) which are queried by the GTS query. The evaluator is invited to answer a set of questions (Table 2) related to the relevance of the retrieved results. For the first question Q1, the evaluator answers "Yes" in case there is at least one result retrieved. Q1 is relevant as for some queries, especially those related to a very recent subject, there is no result retrieved. In case more than one result is retrieved the evaluator answers Q2 as whether the first result is related to the query or not. For instance, the first result for the queries *The Voice* and *San Francisco Weather* were not relevant to the query from Wikipedia. Question Q3 requests the evaluator to choose the most relevant answer amongst the ten first results retrieved. Q4 is dedicated to answer whether Wikipedia webpage contains a table of contents. Q5 requests from the evaluator to choose which image in Yahoo images among the top 10, matches exactly the image in GTS.

Table 2: Experiment questions

Question ID	Question	Possible Answers
Q1	The query retrieved at least one result	Yes / No
Q2	The 1st result is related to the query	Yes / No
Q3	Among the top 10 results, which one is the most related to the query	1 - 10
Q4	Wikipedia's Table of Contents exists	Yes / No
Q5	Which image in Yahoo images among the top 10, matches exactly the image in GTS	1 - 10

7.2 Experiment setup

The results of the experiment for each question are reported in Table 3. The relevance of resources to the queries used shows that YouTube resources are the most relevant to the query. The last column "LO retrieved" of the table represents the percentage of the cases where a learning object has been retrieved according to the question requirement. It represents the conjunctive grouping result of the three websites for each query, i.e. if all the websites score 1 for a given query then "LO retrieved" scores 1 otherwise it is 0.

Table 3: Experiment results

Question	Wikipedia	Youtube	Yahoo Images	LO retrieved
Q1	100%	100%	96%	98%
Q2	92%	100%	94%	90%
Q3-(1^{st})	86%	92%	94%	66%
Q4	88%	--	--	--
Q5-(1^{st})	--	--	12%	--

7.3 Discussion

The experiment results aggregated in Table 3 show that Mole is able to generate a learning object in 98% of the cases (Q1). In terms of relevancy (Q2), 90% of the generated learning objects were related to the query. Q3-(1^{st}) represents the percentage of the cases where the first result retrieved is the most related to the query. This result (66%) denotes the precision of Mole. This means that in 66% of the cases, the resources retrieved by the three websites

[1] http://www.google.com/trends/hottrends

[2] Evaluators are colleagues from the department of information systems. We thank them for their collaboration.

correspond to the best match. Accordingly, in 24% of the cases Mole retrieves a learning object that is relevant to the query but it is not the best match. In 10% of the cases, one of the resources was not related to the query or Mole was not able to retrieve the complete learning object. Although Wikipedia has scored 92% in Q2, in 4% of the cases the table of contents was missed and Mole could not generate the learning web. For Q5-(1^{st}), the score 12% represents the percentage of cases where the first query of Yahoo Images is the image displayed in GTS. This discrepancy between images of the two websites can be explained by the fact that images in GTS are authored by newspapers to which GTS refers. This result does not impact on the relevance of images retrieved by Yahoo Images since 94% of retrieved Yahoo Images correspond to the best 1^{st} match with the queries.

In order to increase the quality of the resources retrieved, future work could focus on three directions to improve Mole results: i) using other web repositories and search engines to retrieve learning resources. Indeed, other semantically focused retrieval systems might provide better performance than the currently used; ii) using CE method systematically to analyze the query and the resources descriptions to help retrieving the best matches. A previous study [11] shows that CE can improve sensibly the precision as it uses methods to analyze the objective of the query and expands the query by using synonyms; iii) using other means to generate the Learning Web. Currently we are using only the table of Contents of Wikipedia while other resources such as domain ontologies, whenever available, might give better navigation options for the learner.

8. Conclusion

In this paper, we argued that hypermedia web resources can be reused to generate e-learning instructional material. This is particularly practical for mobile learning where nomadic users need to learn about real-life situations that they face contextually. Web 2.0 includes many open access repositories, technologies and web services that can be exploited in e-learning applications. The architecture proposed in this paper gathers hypermedia resources from open repositories which are retrieved on-demand, integrated into learning objects, and plugged into the learning web allowing learners to navigate freely through its structure. Learning resources need to be described semantically in order to be re-used and shared. We proposed the use of the Contextual Exploration Method to annotate semantically resources by analysing their descriptions. CE is suitable for unrestricted texts, as it does not require heavy representations and detailed descriptions of linguistic units in the analysis of the language. Mole, the mobile learning system presented generates automatically learning objects and a learning web in response to user queries allowing him/her to learn through sessions. The evaluation performed with Google Hot Trend queries show that Mole was able to generate learning objects relevant to the query in 90% of the cases. While these results are very promising, further research directions have been identified to address some challenging aspects of the approach. One such direction is to further investigate the semantic annotation of resources to increase the system precision. We are currently exploring a system that uses systematically CE to analyse descriptions of resources and also the comments of users. We are also considering improvements to the graphical user interface to make it more intuitive and easy to use for navigation. Besides, more LO layouts are explored to be added to Mole to fit different user profiles and increase the level of customization.

References

[1] di Sciascio, C., Brusilovsky, P., Trattner, C., Veas, E., A Roadmap to User-Controllable Social Exploratory Search, ACM Transactions on Interactive Intelligent Systems, 10(1), 2020, Article 8.

[2] Kevin Burden, Matthew Kearney, Sandra Schuck, Tony Hall, Investigating the use of innovative mobile pedagogies for school-aged students: A systematic literature review, Computers & Education, 138, 2019, pp. 83-100.

[3] Kinshuk, Nian-Shing Chen, I-Ling Cheng, Sie Wai Chew, Evolution is not enough: Revolutionizing Current Learning Environments to Smart Learning Environments, International Artificial Intelligence in Education, 26(2), 2016, pp. 561-581.

[4] Matthew L. Bernack, Jeffrey A. Greene, Helen Crompton, Mobile technology, learning, and achievement: Advances in understanding and measuring the role of mobile technology in education, Contemporary Educational Psychology, 160, 2020, Article 101827.

[5] Leila A. Mills, Gerald Knezek and Ferial Khaddage, Information Seeking, Information Sharing, and going mobile: Three bridges to informal learning, Computers in Human Behavior, 32, 2014, pp. 324–334.

[6] N. Rubens, D. Kaplan, and T. Okamoto. E-Learning 3.0: anyone, anywhere, anytime, and AI. In International Workshop on Social and Personal Computing for Web- Supported Learning Communities (SPeL 2011), Dec 2011.

[7] Jawad Berri, Leveraging Web Information for E-Learning, New Perspectives in Information Systems and Technologies, Volume 2, Advances in Intelligent Systems and Computing, Springer, Volume 276, pp. 477-486, 2014.

[8] Y. Atif, D. G. Sampson, S. Harous, “Supporting the Selection of Digital Resources for Classroom- based Teaching Activities”, In Proc. of the IEEE International Conference on Advanced Learning Technologies (ICALT 2013), Beijing, China, pp. 359-363, Jul 2013

[9] Y. Atif, “Conversational Learning Design in Future Classrooms”, In Proc. of the IEEE International Conference on Advanced Learning Technologies (ICALT 2012), Rome, Italy, pp.178-180, Jul 2012.

[10] Sampson, D., Spector, J. M., Aedo-Cuevas, I., Chen N. S. (Eds.), "Special Issue on Current Advances in Learning Technologies", Educational Technology & Society Journal (ISSN: 1436-4522), vol. 15, no. 4, October (2012).

[11] Dunlap, J. C., & Lowenthal, P. R., Learning, unlearning, and relearning: Using Web 2.0 technologies to support the development of lifelong learning skills. In G. D. Magoulas (Ed.), E-infrastructures and technologies for lifelong learning: Next generation environments. Hershey, PA: IGI Global, 2011. DOI: 10.4018/978-1-61520-983-5.

[12] Seiz-Ortiz, R., Gimeno-Sanz, A., Macario de Siqueira, J., APPRAISALWEB: an online platform for the pedagogical evaluation of Web-based Language Learning Resources, Procedia Social and Behavioral Sciences, vol. 15, pp. 667--671, (2011).

[13] Brusilovsky Peter, Wade Vincent P., Conlan Owen, From Learning Objects to Adaptive Content Services for E-Learning, IGI Global, 2008, pp. 243-262.
[14] Claus Pahl and Edmond Holohan, Applications of Semantic Web Technology to Support Learning Content Development, Interdisciplinary Journal of E-Learning and Learning Objects, 9, 2009, pp. 1-25.
[15] Y. Atif, R. Benlamri, and J. Berri, Learning Objects Based Framework for Self-Adaptive Learning, Education and Information Technologies, Kluwer Academic Publishers, 8(4), 2003, pp. 345–368.
[16] R. Benlamri, J. Berri and Y. Atif, “A Framework for Ontology-aware Instructional Design and Planning”, Journal of E-Learning and Knowledge Society, Vol.2, No.1, pp. 83-96, 2006.
[17] B. Maxim, K. Akingbehin and K. Modesitt, Just-in-Time Learning in Software Engineering, Journal of Computers in Mathematics and Science Teaching, 18(3), 1999, pp. 287-301.
[18] Advanced Distributed Learning – SCORM Version 1.2, 2001, www.adlnet.org
[19] IEEE Learning Technology Standards Committee (LTSC). IEEE P1484.12/D4.0 draft standard for learning object metadata (LOM). IEEE Computer Society, 2002.
[20] A. Zubiaga, V. Fresno, R. Martınez, and A. Pérez Garcıa-Plaza, Harnessing Folksonomies to Produce a Social Classification of Resources, IEEE Transactions on Knowledge and Data Engineering, vol. 25, No. 8, August 2013, pp. 1801- 1813.
[21] J. Hendler, T. Berners-Lee, and E. Miller, Integrating Applications on the Semantic Web, Journal of the Institute of Electrical Engineers of Japan, vol 122, no. 10, pp. 676-680, October 2002.
[22] Resource Description Framework, http://www.w3.org/TR/rdf-concepts/, Accessed 21 February 2014.
[23] J. Berri, D. Maire-Reppert, H.-G. Oh-Jeong, Traitement informatique de la catégorie aspecto-temporelle", T.A Informations, 32(1), 1992, pp. 77-90.
[24] Kim, J.-T., Moldovan I.: Acquisition of Linguistic Patterns for Knowledge-Based Information Extraction. IEEE Transactions on Knowledge and Data Engineering, Vol. 7, No. 5, Oct. 1995 pp. 713-724.
[25] J. Berri, M. Al-Khamis, Information Exploration Using Mobile Agents, WSEAS Transactions on Computers, 3 (3), 706-712, 2004.
[26] Y. Qu and G. Cheng, Falcons concept search: A practical search engine for web ontologies, IEEE Transactions on Systems, Man, and Cybernetics, Part A, 41(4), 2011, pp. 810–816.
[27] Kim Viljanen, Jouni Tuominen, Eetu Mäkelä and Eero Hyvönen, Normalized Access to Ontology Repositories, Proceedings of the Sixth International Conference on Semantic Computing (IEEE ICSC 2012), IEEE Press, Palermo, Italy, September, 2012.
[28] Wikipedia, http://en.wikipedia.org/wiki/Wikipedia, Accessed 28 September 2019.
[29] Mohammed Alzaabi, Jawad Berri, Mohamed Jamal Zemerly, Web-based Architecture for Mobile Learning, International Journal of Infornomics, 3(1), 2010, pp. 213–222.
[30] Yahoo images, http://images.search.yahoo.com/, Accessed 25 February 2020.

Jawad Berri is an associate professor at King Saud University, Saudi Arabia. He received his Ph.D. in Computer Science from Paris-Sorbonne University in France in 1996. Jawad’s research interests focus on context-aware web systems, learning technologies and natural language processing. He has been involved in many projects related to mobile learning, semantic web, automatic summarization, web information filtering and mobile agents for web information discovery. He worked as a researcher at the CNRS – the French National Research Center, the Computer Science Institute at the University of Zurich – Switzerland and Sonatrach – the Algerian Petroleum and Gas Corporation. Jawad’s contributions in research projects in the industry and academia led to the publication of papers in numerous journals and conferences. Jawad is a senior member of the IEEE.

Rachid Benlamri is a Professor of Software Engineering at Lakehead University - Canada. He received his Master’s degree and a Ph.D. in Computer Science from the University of Manchester - UK in 1987 and 1990 respectively. He is the head of the Artificial Intelligence and Data Science Lab at Lakehead University. He supervised over 80 students and postdoctoral fellows. He served as keynote speaker and general chair for many international conferences. Professor Benlamri is a member of the editorial board for many referred international journals. His research interests are in the areas of Artificial Intelligence, Semantic Web, Data Science, Ubiquitous Computing and Knowledge Engineering

Yacine Atif is a professor in informatics at the school of informatics, University of Skövde, Sweden. He received the PhD degree in Computer Science from Hong Kong University of Science and Technology (HKUST) in 1996. After graduation, he worked at Purdue University in the USA as a Post-Doc and then joined a faculty position at Nanyang Technological University (NTU) in Singapore. From 1999 till 2016 he was at UAE University as faculty, then Program Chair at the College of Information Technology. Dr. Atif has made a number of research contributions particularly in the areas of Learning Technology and Social Web. He is also involved in the Technical Programs of several research forums.

Hajar Khallouki is a postdoctoral fellow at the department of Software Engineering at Lakehead University. She received a Master’s degree from the University of Hassan II – Casablanca in 2012, and a Ph.D. in Computer Science from the University of Hassan I – Settat Morocco in 2019. Her research interests are in Context-aware Computing, Ubiquitous Computing and Healthcare Informatics.