

Towards NLP-based Semi-automatic Preparation of Content for Language Learning using LingoSnacks m-Learning Platform

Abdelkarim Erradi, Hind Almerexhi, Sajeda Nahia

Department of Computer Science and Engineering

Qatar University, Doha, Qatar

{erradi, 200752261, 200751881}@qu.edu.qa

Abstract— Vocabulary growth is an important element for language learning but it requires repeated and varied exposure to the new words and their usage in different context. However preparing suitable learning content for effective language learning remains a challenging and time-consuming task. This paper reports the experience of designing and developing a m-Learning platform (named LingoSnacks) for semi-automatic preparation of content for language learning using Natural Language Processing (NLP) services. LingoSnacks Authoring Tools provide an environment of assisted authoring of learning content and delivering it to the learner in game-like interactive learning activities. Empirical testing results from teachers who used LingoSnacks indicate that the participants were able to ease their lessons preparation tasks. Also the resulting learning packages helped learners in vocabulary acquisition as the number of new vocabulary that they can recognize, recall and retain was significantly higher than participants who just used conventional lessons in a classroom.

Keywords- Mobile Assisted Language Learning (MALL); LingoSnacks; Content Authoring for MALL; NLP for Language Learning

I. INTRODUCTION

Interest in learning Arabic is on the rise. However, learning it is a difficult task that involves significant and repeated effort to build enough vocabulary of high-frequency words to be able to engage in meaningful conversations. Moreover, traditional lecturing/teaching methods based on memorization and outdated resources are becoming ineffective. Hence language learning can become difficult and uninteresting. Also there is a lack of affordable high quality and readily available learning resources. While the way people learn has changed dramatically, the traditional language learning methods did not evolve at the same pace. Today's students have grown up completely immersed in technology using mobile gadgets, computers, internet and Facebook: these are the ways they learn and interact with their world. They need a learning experience made for the way they learn. This new generation of learners needs a new kind of study aids that can help them engage and enjoy the learning experience. English learners have several systems like SpellingCity.com and dynamo.dictionary.com [2, 3, 4, 5] available at their reach, yet; that is not the case for Arabic.

Increasing affordability and advanced features of mobile devices have extended their role from communication tools to socialization, entertainment and learning. Internet-enabled mobile devices are increasingly used for enhancing the learning experience particularly for language learning. They can be exploited to allow learners to access and interact with multimedia learning resources from anywhere and at any time while offering an enhanced user experience [1]. With that in mind, Mobile Learning (mLearning) can be considered as a new educational paradigm that can guarantee a useful and continuous language learning practice [9]. For this reason, Mobile Assisted Language Learning (MALL) applications were continuously developed for over a decade to support language acquisition skills [9].

In our LingoSnacks learning platform, we advocate a micro-learning methodology: short learning activities through interaction with micro-content objects for short sessions that can be chained into significant learning programs. LingoSnacks also aims to ease the preparation of learning content by leveraging a range of Arabic Natural Language Processing (NLP) services that can perform several tasks such as morphology analysis, searching dictionaries and other services. Many existing and popular Arabic Natural Language Processing (NLP) [6] tools were extended for the implementation of the NLP services.

LingoSnacks offers several benefits and holds the potential to considerably support Arabic learners by providing a community for teachers to share and reuse content that can be updated, retrieved and exchanged seamlessly. The posted content is used by LingoSnacks mobile application offering game-like learning activities enriched with interactive multi-media content such as text, photos and videos.

II. LINGOSNACKS OVERVIEW

LingoSnacks is a mLearning framework for preparing and delivering learning content to ease Arabic language learning using game-based micro-learning approach for language vocabulary acquisition. The design and evaluation of the learner components of LingoSnacks were reported in a [7] and [8]. This paper focus on the design and evaluation of the teacher components. As shown in Figure 1, LingoSnacks is comprised of three major elements.

The first element is the LingoSnacks Authoring Tool used for semi-automated Learning Content authoring. The tool provides access to relevant Natural language resources and services (such as Dictionary lookup, Stemming, Parts of speech tagging, and Text2Speech services) to ease creating rich learning content. The Authoring Tool allows the content creator to author, publish and

share the content with others by providing a list of words with associated definitions and example sentences, assign appropriate multimedia illustrations to the words and sentences. The author can also attach other related linguistic information such parts of speech for the sample sentences. The learning content is organized into packages that are pushed to the LingoSnacks content server. The learner can explore available learning packages, download desired ones and engage in a range interactive learning activities and games based on the content of the selected package.

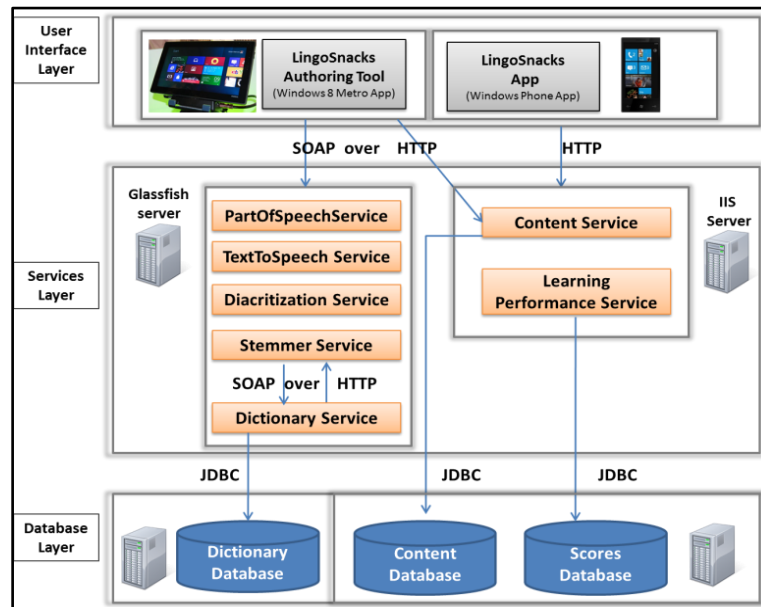


Figure 1. LingoSnacks Deployment Diagram

The second element is the Learning Delivery platform that manages the learning content and learners' performance. This is a cloud hosted service that manages the Learning Packages published from the authoring tool. It also manages the delivery of the learning content to the Learners.

The third element is LingoSnacks Mobile App provides interactive Game-based learning activities to enable learners to acquire new linguistic skills in playful manner. Currently we are targeting modern devices with multitouch technologies such Android and iOS devices. The Mobile App allows searching and downloading learning packages and keeping them up-to-date whenever a new version is available. Also the mobile app allows the users to evaluate learning packages. The app also tracks the learner's performance.

The significant features of LingoSnacks can be summarized as to:

- Allow content creators to semi-automatically author learning content and publish it to learners and/or share it with other users. Basically, the content author provides a list of words to be studied. Then the system suggests the word definition(s) and sample sentence(s) to help learn the spelling and enrich the student's vocabulary. The sentences generation uses a range of machine-readable dictionaries and Arabic corpuses [6]. To enrich the learning content, the content creator can also get further help from the range of Natural Language Processing (NLP) services provided by the platform such as stemming and parts of speech tagging.
- The content provided by the content author in a semi-automated way is the input for the learning activities/games discussed below. The learning content is arranged into packages that are stored in a searchable repository to ease delivery to students. The learning packages can be classified by age-groups/grade levels (e.g., Elementary School, Middle School, High School) and by academic subjects (e.g., science, history, business).
- Allow the learner to select and download the desired learning packages containing set of words to practice along with the associated resources (i.e. recorded audio pronunciation, illustrative photos, and part of speech tags). Then, according to the downloaded content, the system can offer a variety of learning activities/games to practice spelling and vocabulary acquisition. As for the major learning activities, they can be described as follows:
 - Sentence Unscramble: the system scrambles the words of a sentence and the user has to re-order them to build a correct sentence.
 - MatchIt sentences: the system will provide sentences with missing words and the user has to match the word with the correct sentence in order to fill-in the blank.

- MatchIt definitions: similar to the MatchIt sentences activity, but here instead of filling the blank the user has to match between a word and its corresponding definition.
- Which Parts of Speech: the system will offer a complete sentence and several parts of speech tags, the user then will have to select a word in the sentence and immediately select it's corresponding part of speech tag, if the choice is correct both the word and its part of speech will disappear.

Several learning activities in LingoSnacks are inspired from SpellingCity.com but the current focus of LingoSnacks is Arabic. Moreover the system is designed to allow adding more learning activities/games. Although the existing prototype focuses on Arabic, the platform is generic can easily support other languages due to easy portability and extensibility.

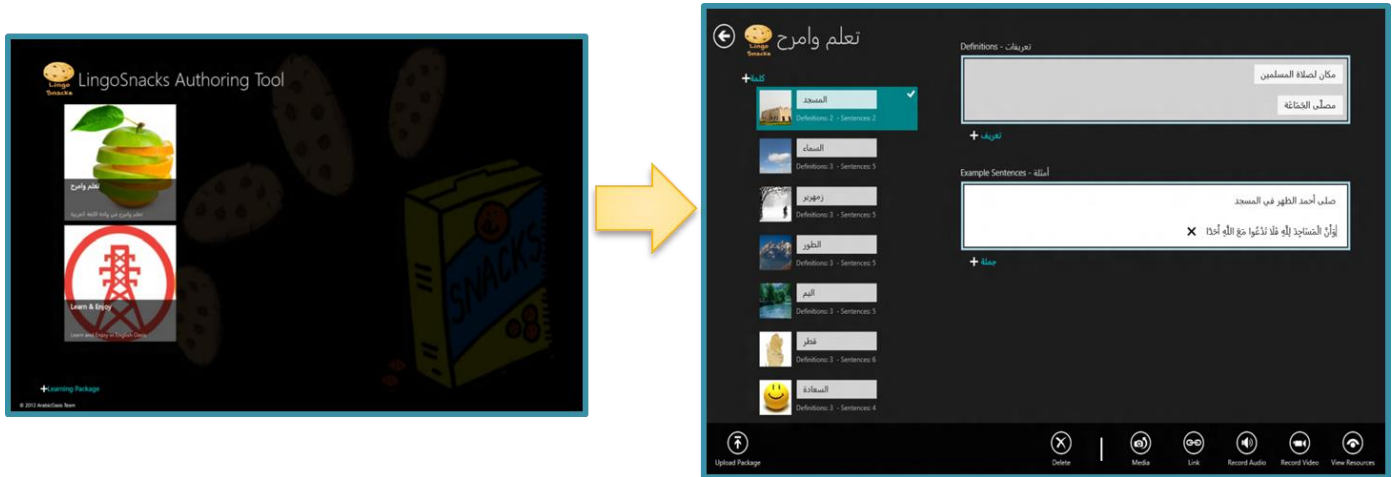


Figure 2. Navigation Design of the User Interface of LingoSnacks Authoring Tool

LingoSnacks uses a Layered Architecture Pattern. The solution is decomposed into three layers with unique roles, so we have the UI Layer, the Services Layer and the Resources Layer.

(1) *The UI Layer includes:* LingoSnacks Authoring Tool which is used by the content creator to author and publish the learning package. The Learning Package Editor is developed as a Windows 8 Metro style application.

And The LingoSnacks App allows the learner to explore available learning packages, download desired ones and engage in a range interactive learning activities/games. The current LingoSnacks App prototype is developed using Silverlight for Windows Phone 7 but porting to other platforms is planned.

(2) *Service Layer:* Various NLP Cloud Services are provided to ease the creation of content. For example the sentences generation uses a range of machine-readable dictionaries and language corpuses. To enrich the learning content, the content creator can also get further help from the range of NLP services provided by the platform such as stemming, part of speech tagging and speech synthesis. These services are implementation using Java EE and rely on several open source implementations particularly Alkhalil Arabic Morphology System [18].

(3) *Resources Layer includes:* Data used by our LingoSnacks system, such as the details of available Learning Package, are stored and managed by SQL server database.

The layered architecture was selected to achieve the following benefits:

- **Availability:** Servers in each layer can be duplicated, so that if one fails, others remain available. For example if the content author requests one of the NLP services, if one instance of these services was unavailable then he can be, without his knowledge, redirected to a replica copy that can satisfy the request.
- **Modifiability:** Separation of concerns enhances modifiability, as the UI Layer, Services and Resource Layer are all clearly encapsulated. Each can have its internal logic modified in many cases without changes rippling into other layers.
- **Performance:** The layered architecture has proven high performance.
- **Scalability:** As servers in each layer can be replicated the architecture scales well. In practice, the data management tier often becomes a bottleneck on the capacity of a system.

Some of the key issues to consider are the speed of connections between the layers and the amount of data that is transferred. Hence, it makes sense to minimize the calls needed between layers to fulfill each request.

III. LINGOSNACKS NLP SERVICES

The key feature of LingoSnacks platform is to allow teachers to semi-automatically author learning content and publish it to students and/or share it with other users. Basically, the teacher provides a list of words to be studied. Then the system suggests the

word definition(s) and sample sentence(s) to help learn the spelling and enrich the student's vocabulary. The sentences generation will use a range of machine-readable dictionaries and Arabic corpuses [6, 31, 32]. To enrich the learning content, the teacher can also get further help from the range of NLP services provided by the platform.

LingoSnacks platform relies on seven NLP services for authoring learning content: StemmerService, DictionaryService, DiacritizationService, PartOfSpeechService, ContentService, LearningPerformanceService and TextToSpeechService. Their interfaces are shown in Figure 3.

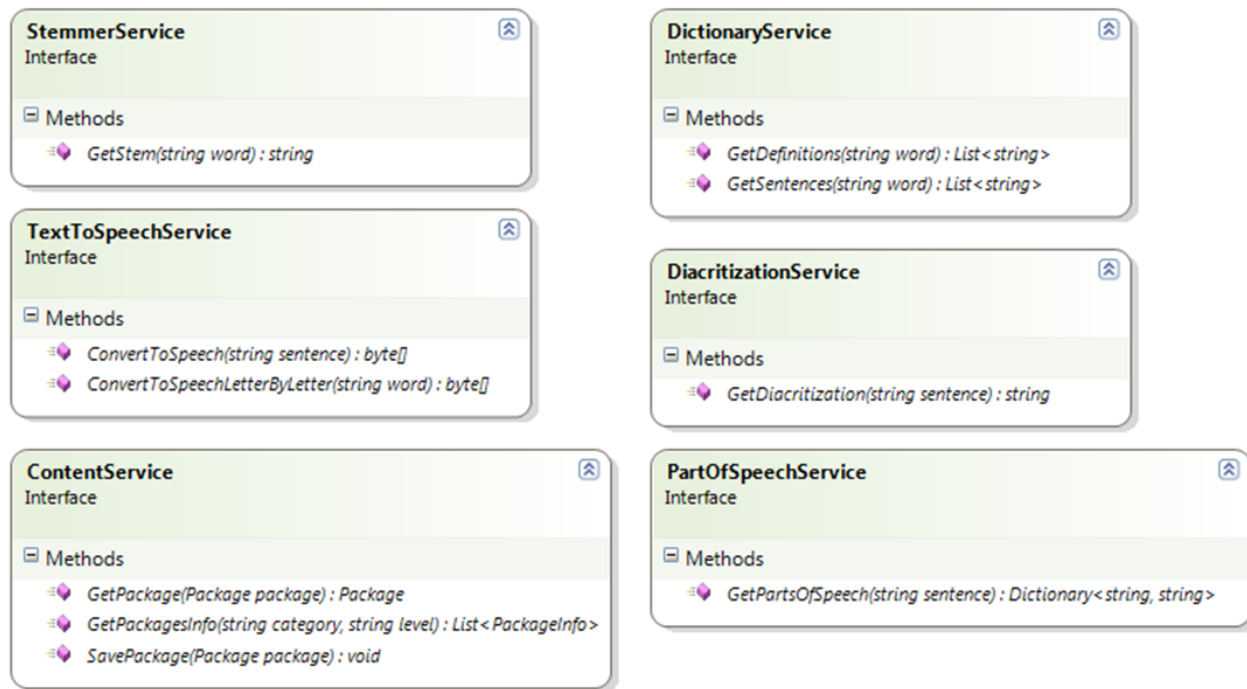


Figure 3. Interfaces of key NLP services

The following sections provide a brief description of the NLP services and how they are implemented.

Stemmer Service: The Stemmer Service is used to reduce a word to its root or stem. The purpose of doing that is to use the obtained root in other services such as looking up the meaning of the word in the dictionary [25]. The implementation of this service relies on Alkhalil Arabic Morphology System [18] open source tool. The merits of this tool is that it not only shows the roots of the words, it also diacritize it and shows many other grammatical analysis information. For further usage, such valuable information can be extracted and incorporated in other services that the instructor may need.

Dictionary Service: This service allows a lookup of an Arabic to Arabic dictionary to provide the meaning any Arabic word. The most appropriate dictionary we have found is the Interactive Arabic Dictionary [17] however it does not provide the promised programmatic API yet. We might consider accessing it using screen scaring techniques alternatively we can explore other dictionaries such as Arabic WordNet [26].

Text to Speech Service: The Text-To-Speech Service is used to convert written text into voices to give the user the ability to hear the accurate pronunciation of Arabic written text. For the implementation of this service we will use Acapela SDK [27]. Acapela offers several Development Kits for a diversity of platforms. Arabic audio in Acapela comes in four different male and female voices, with clear pronunciation and the ability to download and playback audio files in many formats. The text-to-speech service is used to auto-read the generated sentences. The teacher can override the generated speech by recording their own audio pronunciation.

Diacritization Service: When it comes to the Arabic language, there are some characteristics that highly distinguish between it and other languages; one of these traits is the diacritization, which means adding diacritics to words to show how they are spelled or pronounced. In Arabic, it is crucial to diacritize words for better understanding, since many words are spelled the same but hold different meaning, here comes the use of the Diacritization Service as an aid for the Text to Speech Service, because for sentences to be read or pronounced in Arabic, they must be properly diacritized based on each word's position in the sentence [28]. This service is implemented based on Alkhalil Arabic Morphology System [18]. The good thing about Alkhalil tool is that it provides all diacritization possibilities for a given Arabic sentence. Diacritization adds disambiguating short-vowel diacritics to the generated sentences. The teacher can fix the generated diacritization.

Parts of Speech Service: The part-of-speech service assigns a lexical class marker (e.g., verb, subject, object) to each word in a sentence. This service is implemented based on Alkhalil Arabic Morphology System [18]. The latter provides the part of speech possibilities for a given input sentence. Part of speech tagging service is used for tagging sample sentences with the ability of the teacher to override the generated results.

The content provided by the teacher in a semi-automated way is the input for the learning activities/games provided by LingoSnacks mobile application. The authored learning content is organized into packages that are stored in a searchable repository to ease delivery to students using Internet and mobile technologies. The instructor can also add photos and videos to the Learning Package and associate them with particular words. The learning packages can be classified by age-groups/grade levels (e.g., Elementary School, Middle School, High School) and by academic subjects (e.g., science, history, business). The Learning Package key entities are shown in Figure 4.

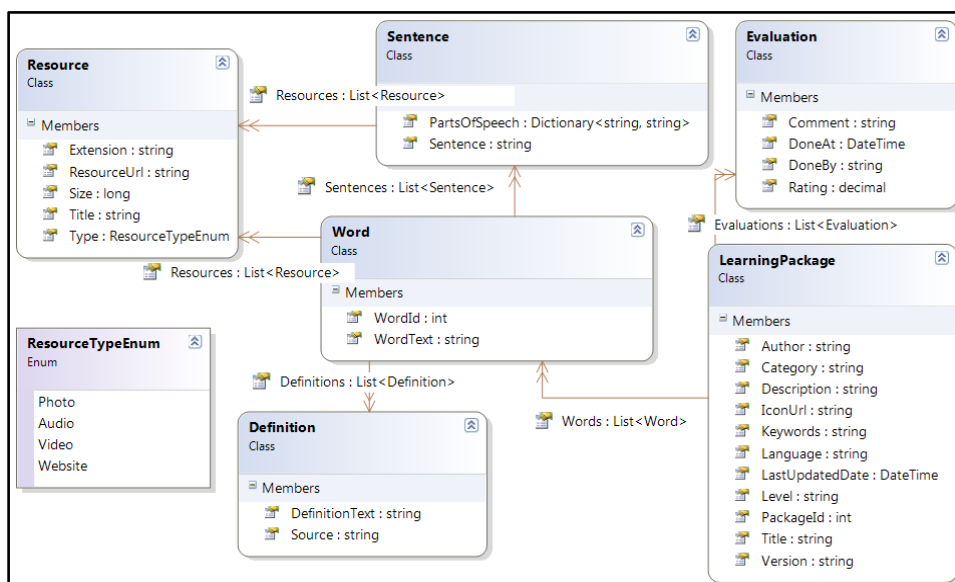


Figure 4. Learning Package key entities

The process for authoring and publishing a Learning Package is depicted in the state diagram shown in Figure 5. The diagram shows the Learning Package possible states and the events that can occur when a learning package is Under Preparation, Submitted for Review, Under Review, Returned, Active or Inactive.

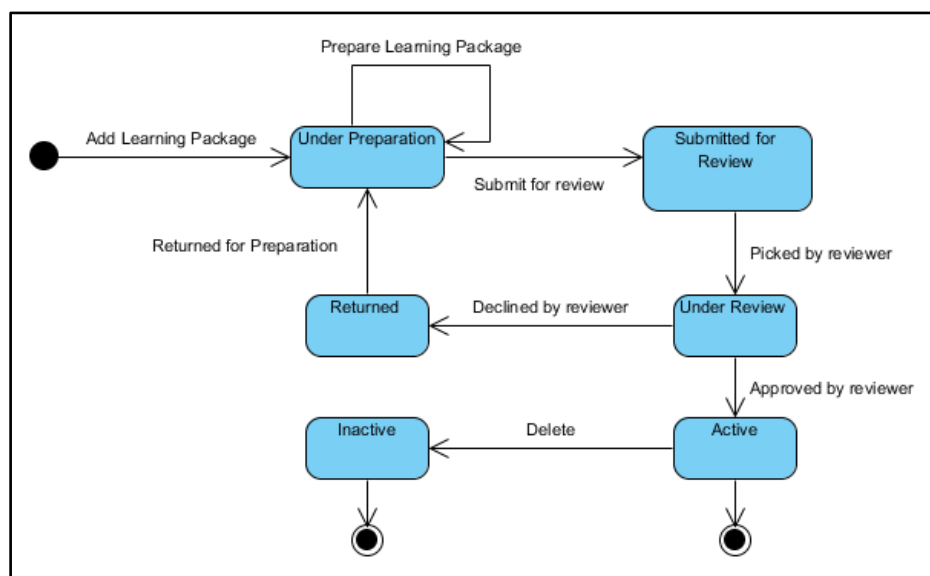


Figure 5 Learning Package State Diagram

Figure 6 shows the System Sequence Diagram for the ‘Get word definitions’ use case. To achieve this use case, the instructor requests the word definitions from the Dictionary Service. The latter uses the Stemmer Service to get the stem of the word before doing a dictionary lookup. A proxy pattern is used to hide the communication details with the remote services. In order to get the list of example sentences from the Dictionary Service, the Stemmer Service gets the stems of words first, then the Dictionary Service communicates with other services such as Al-Waseet Dictionary or the Quran Lookup Service to get example sentences. The Dictionary Service acts as a façade to other services, while the proxy pattern is used as well to hide the communication details with the Stemmer Service.

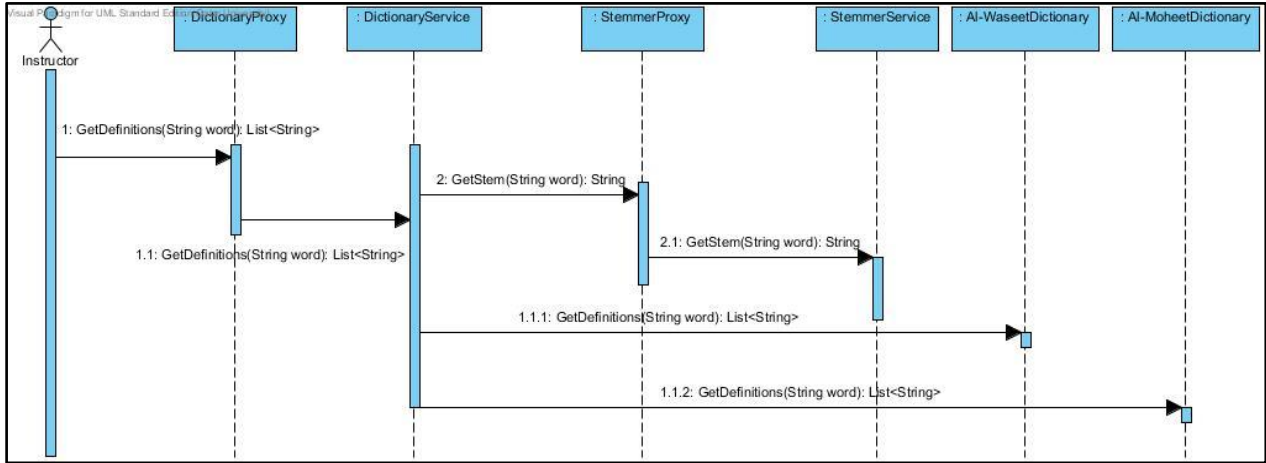


Figure 6. Get word definition Sequence diagram

IV. EVALUATION

As shown in Figure 2, we developed a Windows 8 version of LingoSnacks. In order to evaluate the usability of the prototype and gather users' feedback, we have used a simplified questionnaire using a five-point Likert Scale (1= Strongly disagree up to 5= Strongly agree). We asked 12 kids between the ages of 9 to 12 years to use the tool for a period of time for vocabulary learning. None of 12 students had used a mobile assisted language learning before. As shown in Figure 7, most of the participants liked the application and tried all the activities. They enjoyed using them and they expressed interested to continue using LingoSnacks. Only two girls were not satisfied because they did not like the colors used in the application. We plan to enhance our mobile application to support themes so that the user can change the color scheme used by the application. The students were assisted to answer a usability questionnaire. The key results are shown in Table 1. The evaluation revealed that the perceived usefulness of LingoSnacks is high and helped boost interest in learning. The majority of participants had a positive attitude towards mobile learning as a complement to their formal learning experience in class. They showed enthusiasm to continue self-learning using LingoSnacks. The participants were satisfied with the system and they would like to use it again.

In another experiment, grade 4 students from the same Arabic class were divided into two groups: one group only used traditional school classes and associated practice homework. The other group used LingoSnacks beside their school classes for 3 weeks. Their teacher made available learning packages that are related to their lessons so that the content used in the learning activities is according to their lessons taken in class. After this, both groups of students were tested to assess their learning outcomes. The results indicated positive support for vocabulary acquisition. LingoSnacks users were able to increase their rate of vocabulary acquisition as the number of new vocabulary that they can recognize and recall was significantly greater than participants who just used conventional lessons in a classroom. Additionally, participants exhibited higher motivation to learn since the system is made up of easy-to-use game-like activities that attracts the learners with multimedia content that they can enjoy. This is a significant observation since motivation is an important ingredient in the learning process.

For the usability evaluation of the LingoSnacks authoring tool, we asked 10 people to voluntarily use the system to prepare some learning content. 7 of those people were Arabic teachers and were interested in Arabic learning. 80% of the users were satisfied and they liked the tool and the options available for them. Within a short period, they were able to learn how to use it by themselves without any help or prior training. They were very interested to use the tool because they felt that it gives them an easy way to prepare their own learning content. Two testers were not satisfied much with the authoring tool. They indicated that this will make the learning process more complicated. They prefer the learning applications with ready, pre-configured content.

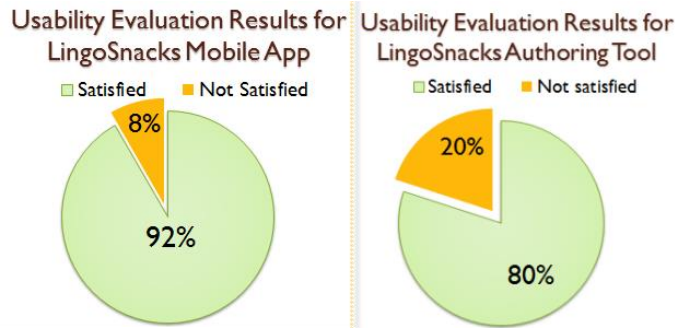


Figure 7. Usability Evaluation Results

Table 1. Participants Feedback on LingoSnacks Usability

Question	Avg. Rating
1. Was navigating (moving) through the learning activities easy to understand and use?	4
3. Was the screen layout easy to follow?	4.5
4. Was the screen layout consistent throughout?	4.5
5. Overall, do you think the system was easy to use?	4.5
6. Would you like to use the system again?	4
7. Did learning using LingoSnacks increased the quality of your learning experience?	5

V. RELATED WORK

Mobile Assisted Language Learning (MALL) is an active area of research and development [1]. The use of smart mobile devices has a major impact on how learning takes place in many disciplines and contexts, including language learning. Many solutions to support design, authoring and delivery of learning activities have been proposed. This section highlights the key related work and emphasizes the differences and similarities between our approach and the available similar works and tools.

SpellingCity.com [2] is a website that offers interactive games and activities to learn English spelling and vocabulary. It provides voice instructions to help the learner to go with each activity. Furthermore, an immediate feedback is returned to the learner containing detailed explanation after each activity. Dynamo.dictionary.com [3] is another eLearning system for vocabulary acquisition. These two systems are similar to our system LingoSnacks except that our system focuses on Arabic language learning. One their major limitation is that they rely on Flash-based implementation which is not compatible with some widely used mobile devices. Additionally, the learning content in our approach is prepared by the instructor and some of it is semi-automatically generated using a collection Arabic NLP services. Another important difference is that our system targets smart mobile devices.

The Ubiquitous English Vocabulary Learning (UEVL) system [5] is a system that is developed to help students to experience the Systematic Vocabulary Learning (SVL) process proposed by Brown and Payne in ubiquitous learning contexts. This SVL process mainly contains five phases: encountering, getting, understanding, consolidating, and using. The UEVL system is constructed using ubiquitous technology and near-synonyms and similar-looking (NSSL) technology, in addition to the use of video clips as learning material to provide the students visual real-world contexts. The ubiquitous technology can sense the situation of students and provide the appropriate learning material for them in real contexts, while the NSSL technology is used to find and define similar-looking words to increase students' awareness of such words. This system has some similar points to LingoSnacks where it is developed to support learning vocabulary anytime and anywhere, and it uses videos as a learning aid. However it does not provide interactive learning activities for students to be engaged in as we have in our system.

VI. DISCUSSION AND CONCLUSION

This paper presented a mLearning system called LingoSnacks that was designed and implemented to help in delivering and accessing the learning content particularly for Arabic language learning. LingoSnacks system allows short sessions of microlearning during any brief opportunities that the student can have in the day. It also addressed the problem of dry content and boring traditional way of learning Arabic because it was designed and implemented as a solution to provide novel and enjoyable ways to learn Arabic using the multi-touch interactive games.

Additionally, the platform facilitates authoring, sharing and exchange of learning content, where the instructor can prepare the learning content and the learner can use it in the different learning activities such as MatchIt sentence, MatchIt definition, Sentence Unscramble game, Parts of speech game, and the Flash cards.

Our system has the potential to change the Arabic learning experience from a boring activity that we have to do to an enjoyable learning experience that the learner would enjoy doing. What LingoSnack does is that it packs the dry Arabic learning content and wraps it in game-like interesting learning activities enriched by relevant multimedia content. So, at the end, the process and the attitude towards learning the language will be positively changed.

For future work, we will continue the system implementation, testing and usability evaluation. The system could also be enhanced to allow end-users to contribute content such as adding new words and definitions as well as relevant multimedia content. We also plan to explore adaptation models and mechanisms for adaptive and personalized learning to tailor the learning package to suit the learner situation based on their educational objectives and preferences, contextual information about the location and circumstances of the learner, and the computing and connectivity capabilities of the device.

ACKNOWLEDGEMENT

This publication was made possible by a grant from the Qatar National Research Fund, award number UREP11-064-1-017. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the Qatar National Research Fund.

REFERENCES

- [1] Bruck, P. A., Motiwalla, L., & Foerster, F. "Mobile Learning with Micro-content: A Framework and Evaluation", (2012). BLED 2012 Proceedings. <http://aisel.aisnet.org/bled2012/2/>
- [2] SpellingCity <http://www.spellingcity.com/> [Accessed: 1 December 2012].
- [3] WordDynamo <http://dynamo.dictionary.com/> [Accessed: 1 December 2012].
- [4] R. Romero, T. Zarraonandia, I. Aedo, and P. Díaz, "Designing usable educational material for English courses supported by mobile devices," in Proceedings of the 6th international conference on HCI in work and learning, life and leisure: workgroup human-computer interaction and usability engineering, Berlin, Heidelberg, 2010, pp. 373–383.
- [5] Huang, Y.-M., Huang, Y.-M., Huang, S.-, Lin, Y.-T. A ubiquitous English vocabulary learning system: evidence of active/passive attitudes vs. usefulness/ease-of-use, Computers & Education (2011), vol. 58, no.1, pp. 273–282.
- [6] N. Y. Habash, "Introduction to Arabic Natural Language Processing," Synthesis Lectures on Human Language Technologies, vol. 3, no. 1, pp. 1-187, Jan. 2010.
- [7] A. Erradi, S. Nahia, H. Almerekhi, L. Al-kailani, "ArabicTutor: a Multimedia m-Learning Platform for Learning Arabic Spelling and Vocabulary", International Conference on Multimedia Computing and Systems (ICMCS 2012), Tangier Morocco, pp. 833 – 838
- [8] A. Erradi, S. Nahia, H. Almerekhi, L. Al-kailani, "LingoSnacks: m-Learning platform for language learning", Colloquium in Information Science and Technology (CIST 2012), Fez Morocco, pp. 149 - 154
- [9] L. Naismith, P. Lonsdale, G. Vavoula, and M. Sharples, "Literature Review in Mobile Technologies and Learning: Report 11," *Educational Technology*, 2004.
- [10] X. Gu, F. Gu, and J. M. Laffey, "Designing a mobile system for lifelong learning on the move," *Journal of Computer Assisted Learning*, vol. 27, no. 3, pp. 204–215, Jun. 2011.
- [11] [31] Latifa Al-Sulaiti and E. Atwell, "The design of a corpus of contemporary Arabic," *International journal of corpus linguistics*, vol. 11, no. 2, pp. 135-171, 2006.
- [12] [32] "Research." [Online]. Available: <http://www.comp.leeds.ac.uk/eric/latifa/research.htm>. [Accessed: 28-Dec-2011].
- [13] [17] "المعجم العربي التفاعلي-الرئيسية." [Online]. Available: <http://www.almuajam.org/>. [Accessed: 28-Dec-2011].
- [14] [18] "SourceForge.net: Alkhalil Morpho Sys - Project Web Hosting - Open Source Software." [Online]. Available: <http://alkhalil.sourceforge.net/>. [Accessed: 28-Dec-2011].
- [15] [26] Black, W., Elkateb, S., Rodriguez, H, Alkhalifa, M., Vossen, P., Pease, A. and Fellbaum, C., (2006). "Introducing the Arabic WordNet Project", in *Proceedings of the Third International WordNet Conference*, Sojka, Choi, Fellbaum and Vossen eds.
- [16] [26] "AWN Browser." [Online]. Available: <http://www.globalwordnet.org/AWN/AWNBrower.html>. [Accessed: 28-Dec-2011].
- [17] [27] "Text to Speech and Voice Solutions - Acapela Group - text to speech - speech synthesis - voice synthesis - speech solutions - voice solutions." [Online]. Available: <http://www.acapela-group.com/index.html>. [Accessed: 28-Dec-2011].
- [18] [28] "diacritizer, Arabic language, morphology, diacritics, Arabic software, Localization, Microsoft Arabic Windows, Word Processor, desktop publisher." [Online]. Available: <http://aramedia.com/diacritizer.htm>. [Accessed: 28-Dec-2011].
- [19]