Text Modeling In Adaptive Educational Chat Room Based On Madamira Tool

[1]Jehad A. H. Hammad, [2]Mochamad Hariadi, [3]Mauridhi Hery Purnomo
Department of Computer Engineering
Institut Teknologi Sepuluh Nopember (ITS)
Surabaya, Indonesia
E-mail: jhammad35@ hotmail.com, mochar@te.its.ac.id, hery@ te.its.ac.id

[4]Nidal A. M Jabari
Department of Computer
Technical Colleges(Arroub)
Palestine
E-mail: njabari@ptca.edu.ps

***Abstract*— This paper discusses how to enhance the ability of text modeling in Arabic during chat sessions. Hanini and Jabari et al. modeled the text in chat sessions, but there is still a problem when using Arabic , because the Arabic language is very difficult to comprehend, has complex derivative and many ambiguities. This paper enhanced the previous study and added MADAMIRA tool to analyze the Arabic text. Monitoring and modeling has been completed through the text modeling process by evaluating the student expressions within the chat session using MADAMIRA tool and machine learning. MADAMIRA tool enables the modeling process to categorize Arabic text into different categories, which makes it easier to use the levels of the used expressions and discover the importance of the chat session between two peers. The process of the student modeling using MADAMIRA and Machine learning will update the student model which gathers information about the student achievements within the AVCM.**

***Keywords*— Model; Chat; AVCM; Student; MADAMIRA; Text**

I. INTRODUCTION

Educational adaptive learning is considered to be intelligent learning. The process of learning depends mainly on student modeling. Student modeling is a methodology to extract the student's characteristics and evaluation process within the virtual classroom[1]. One crucial component is adaptive peer to peer chat[2]. In[2,3] the virtual classroom modeling and student modeling in the adaptive educational process were discussed. Additionally, this paper will discuss the text modeling in the educational adaptive chat room based on methodology and using the MADAMIRA framework[5]. Text model means extracting the evaluation criteria, which can be used to evaluate the text and expressions the student uses while he is chatting with his/ her peer in the virtual classroom. By succeeding to develop and enhance the student modeling in the adaptive chat room, a new component to the student modeling within the chat room will be added. This will be in addition to the last model, which was related to the time investment during the session by the student [4]. Finally, all these models will be integrated into the AVCM student modeling components[2].

II. BACKGROUND AND THEORITICAL FRAMEWORK

A. AVCM

Jabari et al. constructed a methodology and framework for adaptive virtual classroom[2]. They constructed three main facilities in this framework to serve the intelligent learning. These are the following: 1) Adaptive presentation 2) Adaptive testing 3) Adaptive chat.

In order to construct these three facilities it was necessary to design several models[2].

1. Domain Model (DM): DM describes the construction and sequencing of the course.

2. Student Model (SM): SM describes the characteristics of the student.

3. Activity Model (AM) : AM describes the activities which should be performed by the student during the course.

4. Resource Model (RM): RM describes the resources of the activities.

5. Nodes Selector (NS): NS is a tool for selecting the teaching materials objects.

6. Concept Score Evaluator (CScE): CScE is used to evaluate a student in one concept.

7. Cognitive Style Evaluator (CSsE): CSsE is used to discover the student's cognitive style.

8. Chat Room Interface Adapter (CRIA): CRIA is used to tune the chat room interface according to the student's cognitive style.

9. Peer Evaluation (PE): PE is used to evaluate the peer student in a chat session.

The student modeling has many aspects in the AVCM classroom. For example, two methodologies are used in the presentation and testing process. These are Stereotype and Overlay Models. While in chat sessions, three methodologies are used: Text, time modeling and peer evaluation model. Figure 1 below shows the AVCM architecture.

978-1-5386-0549-3/17/$31.00 ©2017 IEEE

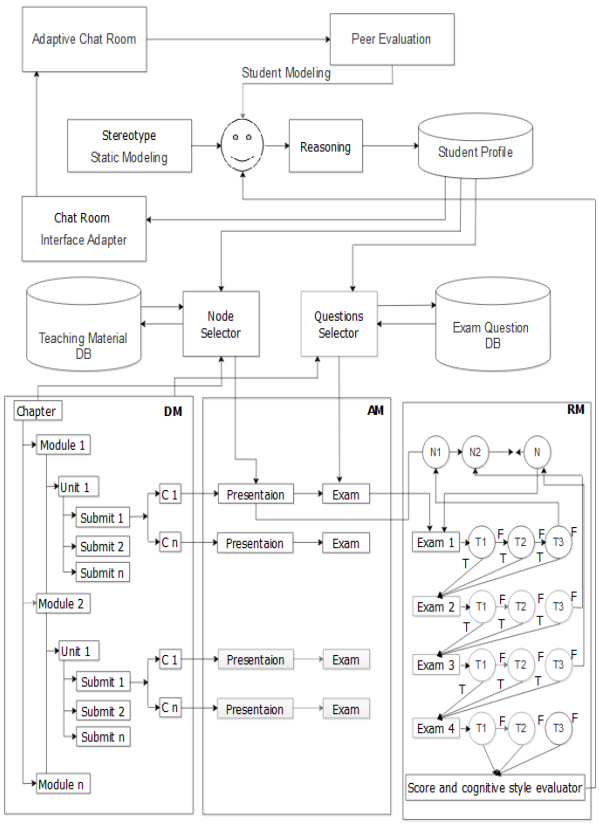

Fig. 1. AVCM architecture.

B. *AVCM Chat Model*

Originally Jabari et al. designed an adaptive chat room, which can adapt its own interface to the student's own characteristics[2].The adaptive chat room is shown in Figure 2.

Fig. 2. AVCM chat model

In[3], the researcher developed a methodology to model the student within the chat room by using three models; text, time and peer model. Each model has its own weight within the overall student modeling.

Chat is a peer-to-peer service for communication with others[6]. A Chat room in AVCM is an additionally separated system within the AVCM[7] that can get advantage from Student modeling in the AVCM. Adaptive chat room is a collaborative tool for methodologies, which can be used to model the students chat whenever they want to discuss with each other. These tools can give evidence about the student knowledge level for each student participating in the chat room while relying on student concept evaluation[2]. The most suitable methodology in our case is that the chat takes input from both student and domain model, and adapts the chat room according to the input score. Different color codes will be used for the different knowledge level. The student who has less knowledge, equal knowledge or higher knowledge than the current student will appear in red color, yellow color or green color respectively. Using this approach it will make it much easier for the current student to decide with whom should he/ she discusses the current concept [2].

C. *Text Modeling In the Adaptive Educational Chat Room*

Hanini and others[3] extracted a mathematical equation to model the student in the adaptive chat room. They continued to model the student by studying each part by itself. They also discussed the text modeling in the adaptive chat room[5].

Jabari et al. discussed a model for observing the content and articulations used by the students during educational chat session with his colleagues about a specific subject of discussion during a specific session[5]. Monitoring has been done through the text modeling process by evaluating the expressions used by the student within the session time. This strategy is used to enhance the general thought of the educational chat among students and make it formalized, which will push the students to care and be more serious during the chat session. Experiments performed as well as previous studies have produced mathematical equations based on the parameters extracted by analyzing and mining the messages submitted by the peers.

To perform the text modeling in the chat room[5], the text, which is used in the chat room, has been divided into eight main groups and another one for unused words called levels. Each level has its own weight in the text final model .The levels are shown in Table 1.

TABLE I.

LEVELS OF CHAT TEXT

Category	Level	Target Tag
Main concept	Level 1	Self
Related directly to the main concept	Level 2	Self
Useful words	Level 3	Self
Positive expressions	Level 4	Both
Agreement expressions	Level 5	Both
Enquiry expressions	Level 6	Both
Respect expressions	Level 7	Self
Negative expressions	Level 8	Peer
Unused expressions	Level 9	Both

The methodology in[5] used to process the text modeling as described in Figure 3.

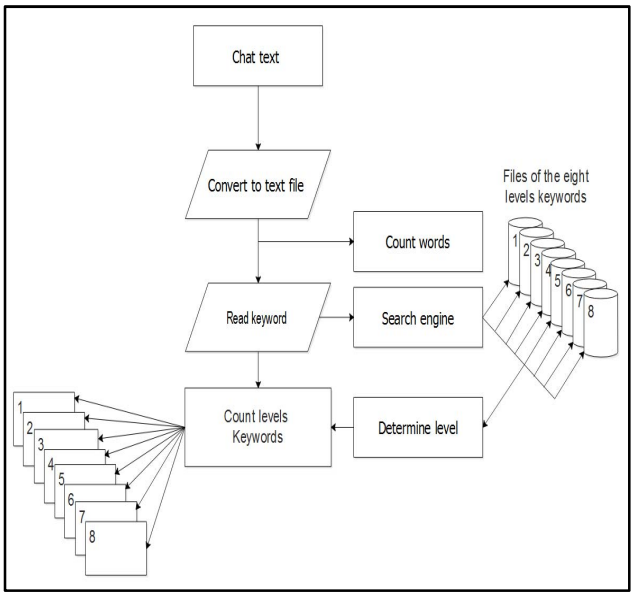

Fig. 3. Text modeling process

And the final text modeling by[5] is shown in Figure 4

Text Model Final Form = -6.91 + (0.9 * Rlevel 1)
+ (0.14 * Rlevel 2)
+ (0.11 * Rlevel 3)
+ (0.06 * Rlevel 4)
+ (0.06* Rlevel 5)
+ (0.18 * Rlevel 6)
+ (0.02 * Rlevel 7)

Fig. 4. Text model form

D. MADAMIRA

MADAMIRA, is a new toolkit used for morphological analysis and disambiguation of Arabic language and vocabulary. This tool uses some of the best aspects of Arabic processing, MADA[11, 8] ; and AMIRA[12]. MADAMIRA uses Java implementation which is much faster than its ancestors (MADA) by more than an order of magnitude, robust, portable, and extensible[9]. This toolkit, which was founded by a team of researchers in Colombia University in 2013, is available as an online tool and it can be downloaded for research purposes.

Arabic is a complex language that has many challenges to NLP, because of the following three main factors:

1: Arabic has a rich morphology system[10].

2: Arabic is highly influenced by its ambiguity, its diacritic-optional writing system and common deviation from spelling standards[9].

3: Due to modern dialects, the language significantly diverges from MSA, which uses the language of the news and formal education. Using NLP tools built for MSA and dialectal Arabic (DA), it is possible to be plagued with very low accuracy.

MADAMIRA has additional components by AMIRA. The Working of MADAMIRA follows the Input text, which goes to the Preprocessor; this Preprocessor cleans the text and changes it to the Buckwalter representation which is used in MADAMIRA. The text is moved towards the Morphological Analysis component, which contains a list of all possible analyses for each word. The text and analyses are then processed to a Feature Modeling component, which uses SVM and language models to predict the word's morphological features. Closed-class features include SVMs, while open-class features of language uses lemma and diacritic forms. The ranking component scores each word's analysis list which is based on analysis accuracy and agrees model predictions, which counts its score. According to the demands and requests from the user, the scoring of each word is passed to the Tokenization segment to get a tweaked tokenization for the word[9].

MADAMIRA can be installed as Standalone mode, or Server-client mode of operation. Figure 5 shows the overview of MADAMIRA architecture[9].

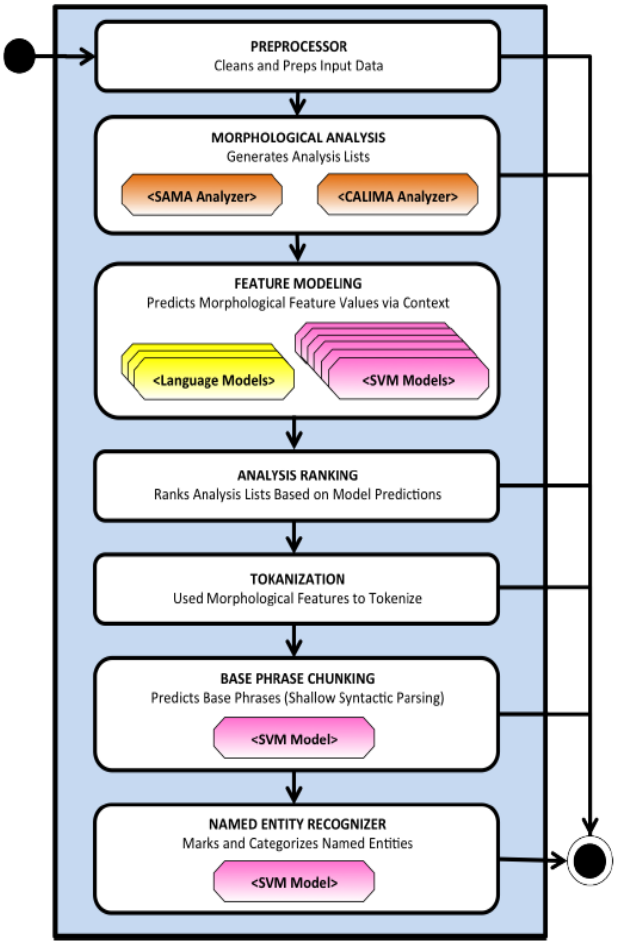

Fig. 5. Overview of the MADAAMIRA architecture

The crucial system assets (analyzers and models) are demonstrated. Input text enters the Preprocessor, and moves through the system, with every part adding extra data that resulting segments can be utilized. Based upon the requested output, the procedure can exit and return results about at various positions in the sequence[9].

MADAAMIRA uses 11 different schemes for tokenizing input text, with individual specifications, while MADA only two methods of tokenizing. An online demo of MADAMIRA, can be accessed by following link: http://nlp.ldeo.columbia.edu/MADAMIRA/.

III. TEXT MODELING IN ADAPTIVE EDUCATIONAL CHAT ROOM BASED ON MADAMIRA

As we have seen in the latest research, many models are created to support the virtual classroom. Student model is one of the main components of AVCM. Student model is gathered from many sources. These are:

1. Stereotype model
2. Overlay model
3. Testing model
4. Chat model
5. Text model
6. Time model
7. Peer model

The student modeling process is described in Figure 6.

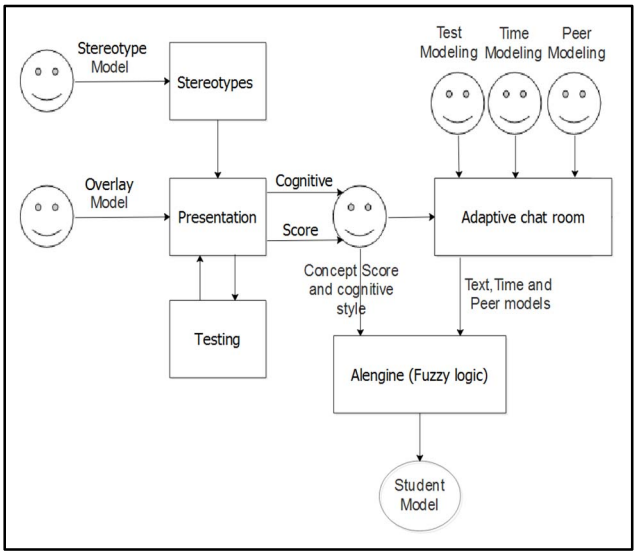

Fig. 6. Student modeling process in AVCM

Back to the text modeling which represents our concentration, we have seen that the latest studies found a methodology for text modeling as described in Figure 4. But the latest studies didn't find the suitable way for Arabic language mining and analysis in order to apply the latest methodology of text modeling.

MADAAMIRA is a solution for Arabic language tokenization for the following reasons:

A) We need only the main words in the text to be used in the text levels discussed in this paper.

B) Levels require that the words should be identified as (verb, nominal, particle and proper noun) which can be achieved using MADAMIRA.

C) MADAMIRA results can be transferred to a text file, which will make it easier to analyze the levels.

Let's take an example for one Arabic chat sentence:

" هل درست المفتاح الاساسي في قواعد البيانات "

This sentence will be analyzed using MADAMIRA engine as the following (see Figure 7).

Fig. 7 MADAMIRA Arabic sentence analysis

As seen in Figure 7, the chat sentence is analyzed as (sentence maker, verb phrase, noun phrase, and propositional phrase). We can now apply the text modeling process described earlier in this paper in table I, as described in table 2:

TABLE II
TEXT LEVELS EXTRACTION USING MADAAMIRA TOOL

Level 1	Level 2	Level 3	Level 4	Level 5	Level 6	Level 7	Level 8	Level 9
1	2	1	0	0	1	0	0	4

When we apply text model as discussed in figure 4, we will get the student model only for this sentence as the following:

Text model = (6.91+(0.9*3)+0.14*1)+(0.11*0)+(0.06*0)+(0.06*1)+(0.02*0)

Text model = 9.75

We can now update the process of the Student model, which was described in Figure 6 to be as described in Figure 8.

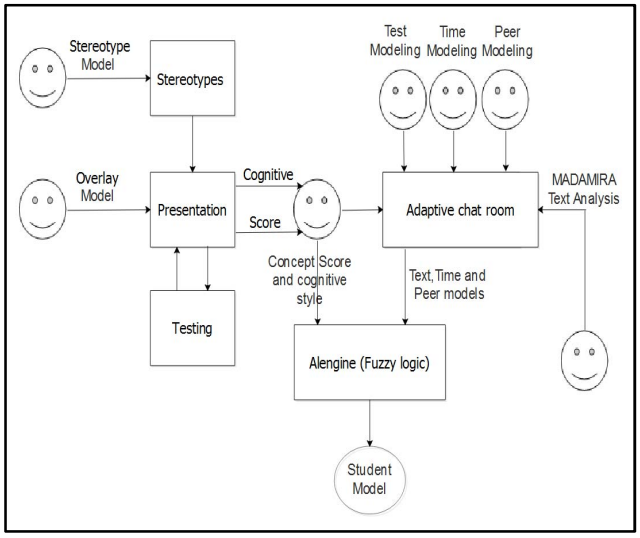

Fig. 8. Text modeling based on MADAAMIRA

IV. OUR PROPOSED STUDENT ADAPTIVE CHATTING MODEL

In the last study by Hanini and Jabari et al. They describe abstract methodology. They show how to extract the nine levels of the keywords in chat sessions as shown in Figure 3.This methodology used only English, and doesn't support the usage of Arabic.

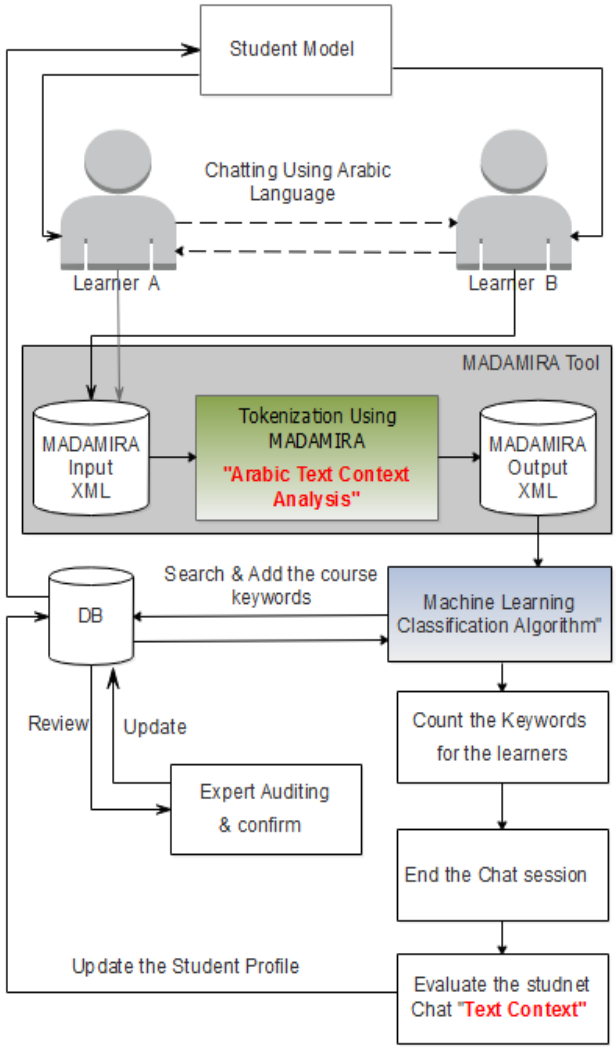

Fig. 9. Text modeling based on MADAAMIR

In our study, we created a new methodology to perform this task as shown in Figure 9.

Our work focused on the following :

A) Analyzes the used Arabic text in the chat room which isn't discussed in the previous studies.

B) Convert the normal educational chatting room into an intelligent one by using a machine learning classification algorithm.

The outcome of our study will be that educational chatting rooms will be smart and will support Arabic, which will be a new achievement that hasn't been accomplished before.

Figure 9 shows the general schema for the text modeling based on MADAMIRA, which is a tool for the analysis of Arabic text.

The system will include a preprocess phase; the aim of this phase is to train the system to make it intelligent by applying many chatting experiments.

The first step of the preprocess stage will be adding the education course parameters and the basic keywords, by the course expert, which will be based on the 9 levels as shown in Table I. These keywords will be the base for the training of the system.

The second step will be the training process; at this stage we'll apply chatting experiments. In each experiment, the learners will be chatting using Arabic. After the session expires the used text will be saved in the MADAMIRA Input XML file. This text will be analyzed and tokenized by MADAMIRA and stored in the MADAMIRA Output XML file. It will then be used by the machine learning classification algorithm. The algorithm will read the keywords at the MADAMIRA output XML file. It will then search the keywords in the nine levels at the database. If the keyword is found, the algorithm will count the keyword for the learner. Nine counters will be used to count the keywords that are used in chat session by each learner. If the keyword hasn't been found in the first eight levels, the algorithm will add it to level nine (Unused words). At the same time, the algorithm will calculate the similarity of the keyword, based on this value the algorithm will decide to add the keyword to the correct level.

In the third step, the course expert will audit the keywords which are added by the system. The aim of this step is to insure that the keywords are added to the correct level.

The algorithm will count the keywords for each learner, and use the values to calculate the text model value using the mathematical formula as shown in Figure 4. Then the system will update the value for the student in the student model.

This experiment will be executed multiple times until we find that the system is learned and almost no new keywords have been found out of the database.

After we finish the system training, the learners can directly chat together without using MADAMIRA, because the system will be intelligent.

V. CONCLUSION

The latest research found a methodology for text modeling in educational chat room. Arabic language used in chat session is a complex language, which needs more tools to be modeled in chat session. Chat expressions as described in previous studies are from several levels and each level should have a rate of occurrences.

MADAMIRA as a tool used in Arabic language tokenization is a suitable tool to solve the problem of analyzing the Arabic text chat and finding the value of the levels to be used in text model.

Using the proposed schema would enable the course Moderator to better model the student's performance during their discussions about course concepts .

References

[1] P. Brusilovsky, "Adaptive hypermedia for education and training," *Adaptive technologies for training and education,* vol. 46, 2012.

[2] N. A. Jabari, M. Hariadi, and M. H. Purnomo, "Intelligent Adaptive Presentation and E-testing System based on User Modeling and Course sequencing in Virtual Classroom," in *IJCA*, 2012, pp. 0975-8887.

[3] M. HANINI, R. TAHBOUB, and N. A. JABARI, "STUDENT MODELING IN ADAPTIVE EDUCATIONAL CHAT ROOM," *Journal of Theoretical & Applied Information Technology,* vol. 58, 2013.

[4] R. Tahboub and M. Hanini, "Time Modeling in Educational Chat Room," *European Journal of Scientific Research,* vol. 122, pp. 27-35, 2014.

[5] M. Hanini, N. A. Jabari, and R. Tahboub, "Text Modeling in Adaptive Educational Chat Room," *International Journal of Computer Applications,* vol. 103, pp. 33-37, 2014.

[6] T. Davey, A. Envall, M. Gernerd, T. Mahomes, M. Monroe, J. Nowak, *et al.*, "Instant Messaging: Functions of a New Communicative Tool," *Anthropology 427: Doing Things with Words,* 2004.

[7] H. Jabari, "adaptive web–based virtual classroom based on student modeling," 2010.

[8] N. Habash, O. Rambow, and R. Roth, "MADA+ TOKAN: A toolkit for Arabic tokenization, diacritization, morphological disambiguation, POS tagging, stemming and lemmatization," in *Proceedings of the 2nd International Conference on Arabic Language Resources and Tools (MEDAR), Cairo, Egypt*, 2009, pp. 102-109.

[9] A. Pasha, M. Al-Badrashiny, M. Diab, A. El Kholy, R. Eskander, N. Habash, *et al.*, "Madamira: A fast, comprehensive tool for morphological analysis and disambiguation of arabic," in *Proceedings of the Language Resources and Evaluation Conference (LREC), Reykjavik, Iceland*, 2014.

[10] N. Y. Habash, O. C. Rambow, and R. M. Roth, "Mada+ tokan manual," 2010.

[11] N. Habash and O. Rambow, "Arabic tokenization, morphological analysis, and part-of-speech tagging in one fell swoop," in *Proceedings of the Conference of American Association for Computational Linguistics*, 2005, pp. 578-580.

[12] M. Diab, N. Habash, O. Rambow, and R. Roth, "LDC Arabic treebanks and associated corpora: Data divisions manual," *arXiv preprint arXiv:1309.5652,* 2013.