Arabic Curriculum Analysis

Hamdy Mubarak*, Shimaa Shoukry, Ahmed Abdelali* and Kareem Darwish***
Qatar Computing Research Institute
Hamad Bin Khalifa University
Doha, Qatar
*{hmubarak,aabdelali,kdarwish}@hbku.edu.qa
Qatar University
Doha, Qatar
**st1103366@student.qu.edu.qa

Abstract

Developing a platform that analyzes the content of curricula can help identify their shortcomings and whether they are tailored to specific desired outcomes. In this paper, we present a system to analyze Arabic curricula and provide insights into their content. It allows users to explore word presence, surface-forms used, as well as contrasting statistics between different countries from which the curricula were selected. Also, it provides a facility to grade text in reference to given grade-level and gives users feedback about the complexity or difficulty of words used in a text.

1 Introduction

Effective language curricula are critical to teaching communication skills that are required in professional and academic settings. Building suitable curricula or updating existing ones is typically a laborious and time consuming process often requiring many specialists working together. Due to its complexity, it is imperative to have tools that ascertain if the curricula achieve the desired learning objectives, as measured for example by vocabulary level. Developing a platform that analyzes curricula can help identify shortcomings and whether they are tailored to desired outcomes. Natural Language Processing (NLP) can provide automated methods to perform such analysis and provide feedback to curricula developers.

A wealth of research devoted to build, curate, and assess educational materials has been published for English and other Latin languages (Tyler, 1950; Oliva, 2005; Braun et al., 2006; Soto, 2015). Though some recent NLP work on Arabic has addressed language learning, readability and textbook assessments (Zaghouani et al., 2014; Zalmout et al., 2016; Al Khalil et al., 2018), the work is limited with rather scarce resources and tools. This paper aims to contribute to curricula assessment, and fill some of the gaps in the literature. We focus on analyzing Arabic curricula taught in Gulf countries at elementary school level. We built a tool that analyzes curricula by providing: statistics about word usage and morphological forms in different grades; words belonging to specific categories, such as food or animals; comparison with other curricula; and complexity levels of words in a text according to selected grades. The tool provides insights into the strengths and weaknesses of curricula and highlights what they cover in terms of vocabulary and morphological constructs. Further, we added some features of potential help to instructors and learners. These include word pronunciation, using text-to-speech, English translation, using machine translation, diacritized forms of words, using automatic diacritization, and linguistic information, such as word segmentation and part-of-speech tagging. As far as we know, this is the first system that: i) allows for browsing and comparing word usages in Arabic curricula from different countries, and ii) showcases whether students in a particular grade would likely understand pieces of text.

2 Related Work

While the research in readability is not novel; advances in technology have permitted researchers to explore further the topic and propose formulations to approximate the difficulty of texts for readers. Benjamin et. al, (2012) surveyed the developments in the field of readability from the perspective of education, linguistics, cognitive science, and psychology, and provided recommendations for the use of

Proceedings of the 27th International Conference on Computational Linguistics, pages 80–86
Barcelona, Spain (Online), December 12, 2020.

such evaluation techniques. Collins-Thompson (2014) explored the challenges for automatic assessment of text readability and highlighted the opportunities for using automatic modeling to predict the reading difficulty of texts. Al-Khalifa and Al-Ajlan (2010) proposed a tool for readability analysis and applied this to curricula in Saudi Arabia. Zalmout et al. (2016) described a process to analyze the textbooks of two different English teaching methods for English as a Second Language (ESL) by using readability scoring technique. Al Khalil et al. (2018) presented an Arabic reading corpus that was collected from textbooks from first to twelfth grade from United Arab Emirates and works of fiction to enhance the inadequate resources that effected educational applications. García Salido et al. (2018) proposed a lexical tool for academic writing in Spanish and described the data extraction from a corpus of academic texts. This tool basically provides insight into how to use typical vocabulary for academic genre in order to build an entire text.

Arabic is a complex language with rich morphology. Stems are typically derived from a set of roots using predefined stem templates. Affixes can be attached to stems to generate words (surface forms). For example, the word وسيكتبونها ("wsyktbwnhA" – "and they will write it")[1] has two prefixes (*and* and *will*) and two suffixes (*they* and *it*). Further, Arabic is typically written without diacritics (or short vowels) which are essential to understand meaning and properly verbalizing words. This increases the complexity when analyzing Arabic texts.

3 Data Collection

We acquired the text versions of the Arabic subject primary school curricular textbooks from six Gulf countries covering grades 1 through 6 from either 2014 or 2015[2]. These countries[3] are: Bahrain (BH), Kuwait (KW), Oman (OM), Qatar (QA), Saudi Arabia (SA), and United Arab Emirates (AE). Statistics are shown in Table 1. Table 2 has example sentences from different grades and shows that the text from grade 1 is direct and simple and is comprised mainly of short declarative sentences. On the other hand, text from grade 6 is more complex, at the vocabulary and sentence structure levels, with longer sentences.

Country	Sentences (K)	Tokens (K)	Unique Lemmas (K)
QA	22	121	<u>6.4</u>
OM	<u>20</u>	<u>110</u>	6.8
KW	29	170	8.1
SA	**40**	**194**	**8.7**
AE	24	134	7.9
BH	31	166	8.1
All	166	895	10.5

Table 1: Corpus statistics for all grades per country. Highest and lowest numbers are written in bold and underlined fonts in order.

Grade	Example
1	دَخَلَ حَمَدُ غُرْفَةَ أُمِه وَهِي مَرِيْضَة. شَاهَدَ فِي الغُرْفَةِ زُجَاجَةَ دَوَاءٍ، فَقَالَ فِي نَفْسِهِ: سَوْفَ أَتَنَاوَلُ هَذَا الدَواءَ Hamad entered his mother's room while she was sick. He saw a bottle of medicine in the room. He said...
6	إِنَ الْإِنْسَانَ الْعَامِلَ يَجِدُ أَنَ الْأَعْمَالَ الْمُفِيدَةَ الَتي يَقُومُ بِهَا تُسْهِمُ فِي صُنْعِ الْحَضَارَةِ الْإِنْسَانِيَة فِي حِينِ... The working man finds that his useful work contributes to the making of human civilization, while ...

Table 2: Sample sentences from QA curriculum (from Grade 1 and Grade 6)

[1] Buckwalter transliteration and translation are provided.

[2] We thank The World Organization for Renaissance of Arabic Language (WORAL) for data collection and preparation.

[3] We use ISO 3166-1 alpha-2 for country codes.

4 System Description

System Architecture: An overview of the system functionalities is illustrated in Figure 1, and the system can be publicly accessed using the following URL: **curriculum.qcri.org**. After the acquisition of the textbooks collection, we used the publicly available Farasa Arabic NLP toolkit to process the text. This includes morphological segmentation (Abdelali et al., 2016), diacritization (Darwish et al., 2017); and lemmatization (Mubarak, 2018). These steps are crucial to enhance the analysis given the complexities of Arabic. Next, language experts classified lemmas into 50 categories (ex: Function Words, Human, Animal, Food, History, Politics, Travel, Religious Acts, etc.)
The system provides the following functions: Term Usage, Category, Statistics, Differences, and Text Grading. It also uses Text to Speech (TTS) , Machine Translation (MT), and Farasa Tools to pronounce, translate, and provide morphological analysis of lexical items respectively.

Design: To implement our tool, we used Django[4], a Python web framework for the rapid development of database-driven websites with high performance web applications. The framework supports model-view-controller (MVC) design patterns to separate the data model and business rules from the user interface. Accordingly, the system modules are separated to ensure reuse and support multiple users and sessions.

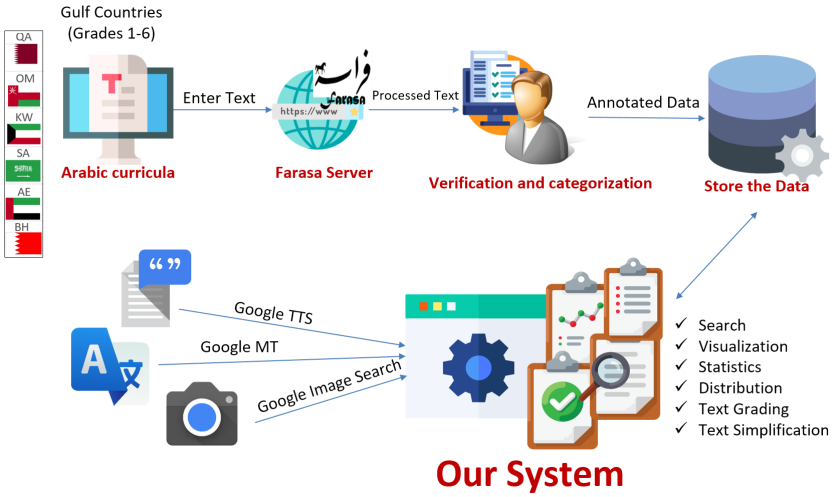


Figure 1: System architecture and functionalities

5 System Functionalities

The system provides five main functions: Term Usage, Category, Statistics, Differences, and Text Grading.

Term Usage: This provides the distribution of input words in all or a subset of grades and countries, and displays all relevant word forms as word clouds as shown in Figures 2, 3 and 4. For ambiguous words, users can select either a diacritized form or an undiacritized form. If the input word is in English, it provides the most frequent translation and displays its information. For any displayed word, users can get translation, listen to pronunciation, and obtain morphological information while hovering (Figure 3).

[4] https://www.djangoproject.com/

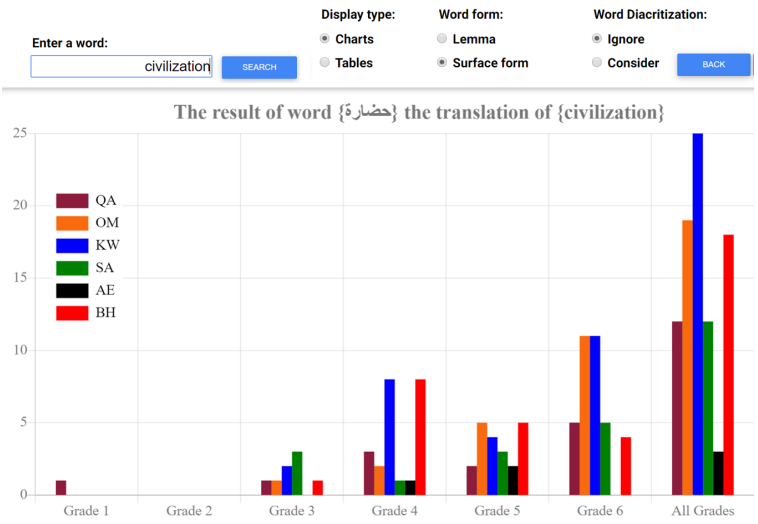


Figure 2: Frequency distributions for the word “civilization” per country and grade

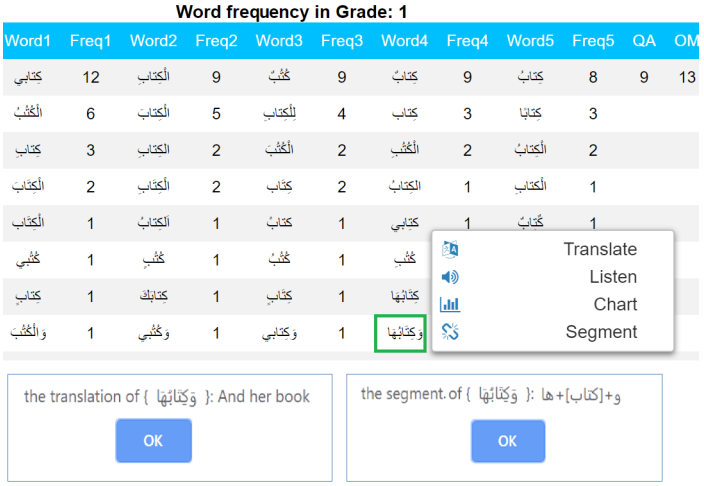
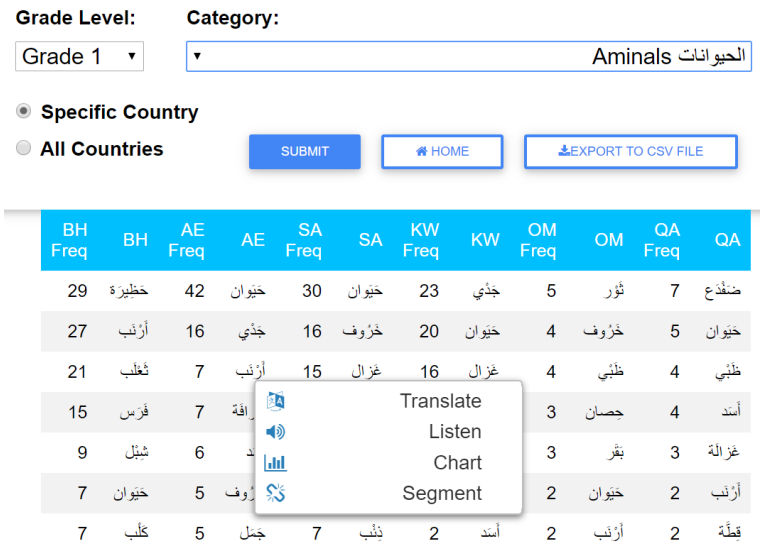


Figure 3: Frequency of word forms for the lemma كتاب (ktAb) “book”, translation and segmentation

Figure 4: Word cloud for surface forms of the lemma استطاع (AstTAE) “he could” in each grade

Category: Users can browse words belonging to a specific category per country and/or grade as shown in Figure 5. Such functionality gives a glimpse into the overall coverage of a given topic/category.



BH Freq	BH	AE Freq	AE	SA Freq	SA	KW Freq	KW	OM Freq	OM	QA Freq	QA
29	حَظِيرَة	42	حيوان	30	حَيَوان	23	جَدْي	5	ثَوْر	7	ضَفْدَع
27	أَرْنَب	16	جَدْي	16	خَرُوف	20	حَيَوان	4	خَرُوف	5	حَيَوان
21	ثَعْلَب	7	أَرْنَب	15	غَزال	16	غَزال	4	ظَبْي	4	ظَبْي
15	فَرَس	7	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	3	حِصان	4	أَسَد
9	شِبْل	6	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	3	بَقَر	3	غَزالَة
7	حَيَوان	5	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	2	حَيَوان	2	أَرْنَب
7	كَلْب	5	جَمَل	7	ذِئْب	2	أَسَد	2	أَرْنَب	2	قِطَّة

Figure 5: Listing for the "Animal" category in grade 1 per country

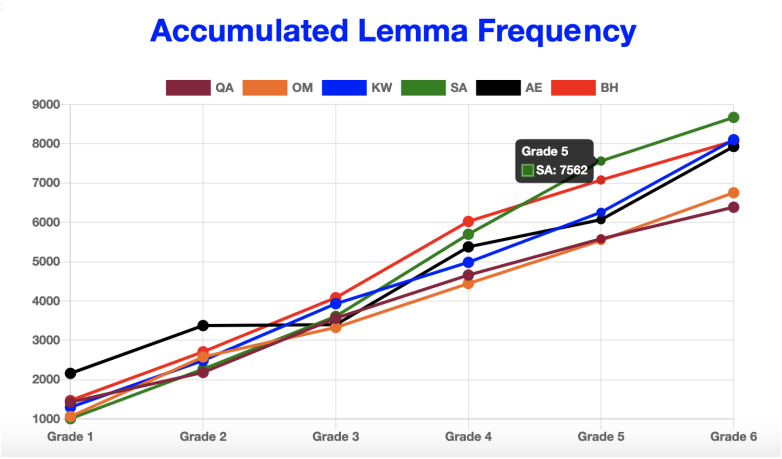


Figure 6: Accumulated lemma frequencies at each grade and for all countries

Statistics: This option shows the distribution of all the lemmas of each country and grade. Results can be shown per grade or accumulated, meaning that for each grade, results of all previous grades are also included. From Figure 6[5] we see that Arab students learn $\sim$ 1.5k lemmas in grade 1 and end up learning 8k lemmas in grade 6. Also, the curricula of QA and OM have lower vocabulary richness compared to the rest of the Gulf countries. This is very important to experts in the field of curriculum development.

Differences: To compare a curriculum of a specific country with that of other countries, users can browse words that are unique to this country in a selected grade as shown in Figure 7.

[5]AE files for grades 3 and 5 are incomplete, and we will solve this in the next release.

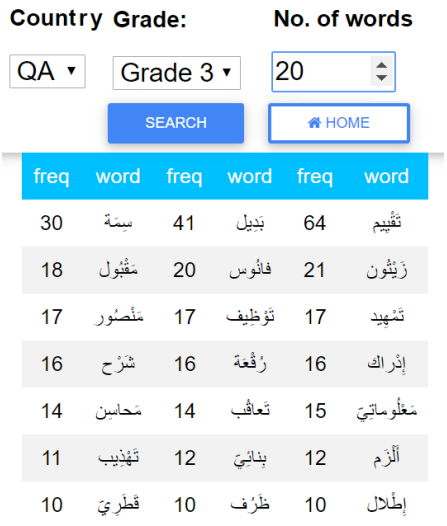


Figure 7: Words that appear uniquely in QA grade 3

Figure 8: Text grading results for grades 1 and 4

Text Grading: This feature allows users to spot "difficult" words in a given input texts. Difficulty of a text can be measured using different methods. The system shows only words whose lemmas did not appear in the selected or preceding grades. As shown in Figure 8, for the input text, the system highlights difficult words in grade 1 (ex: خبراء، جرأة، فلسفة ("xbrA', jr>p, flsfp" – experts, audacity, philosophy)) as they are not seen in the textbooks of the selected grade. When we select higher grades, difficult words decrease. For the second example in Figure 8, the former words in the text are no longer considered difficult for grade 4 except the words ثراء، تحويلية ("tHwylyp, vrA' " – transformative, richness).

6 Conclusion and Future Work

We presented a tool for curricula analysis. For demonstration, we used a collection of elementary school grades (grades 1 to 6) from all Gulf countries. The system provides valuable insights into word usages, vocabulary coverage, and the richness of the curriculum for each grade level. Also, it provides a function for text grading in reference to a grade level, word pronunciation and translation, and morphological analysis. In the future, we aim to extend the collection to other countries and to improve text grading by considering syntactic and semantic features and give grade-level scores for input texts. We are also considering text simplification.

References

Ahmed Abdelali, Kareem Darwish, Nadir Durrani, and Hamdy Mubarak. 2016. Farasa: A fast and furious segmenter for arabic. In *Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: Demonstrations*, pages 11–16.

Hend S Al-Khalifa and Amani A Al-Ajlan. 2010. Automatic readability measurements of the arabic text: An exploratory study. *Arabian Journal for Science and Engineering*, 35(2 C):103–124.

Muhamed Al Khalil, Hind Saddiki, Nizar Habash, and Latifa Alfalasi. 2018. A leveled reading corpus of modern standard arabic. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan, May. European Language Resources Association (ELRA).

Rebekah George Benjamin. 2012. Reconstructing readability: Recent developments and recommendations in the analysis of text difficulty. *Educational Psychology Review*, 24(1):63–88.

Sabine Braun, Kurt Kohn, and J Mukherjee. 2006. *Corpus Technology and Language Pedagogy: New Resources, New Tools, New Methods. English Corpus Linguistics Vol 3*. Lang.

Kevyn Collins-Thompson. 2014. Computational assessment of text readability: A survey of current and future research. *ITL - International Journal of Applied Linguistics*, 165(2):97–135.

Kareem Darwish, Hamdy Mubarak, and Ahmed Abdelali. 2017. Arabic diacritization: Stats, rules, and hacks. In *Proceedings of the Third Arabic Natural Language Processing Workshop*, pages 9–17, Valencia, Spain, April. Association for Computational Linguistics.

Marcos García Salido, Marcos Garcia, Milka Villayandre-Llamazares, and Margarita Alonso-Ramos. 2018. A lexical tool for academic writing in spanish based on expert and novice corpora. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Osaka, Japan, May. European Language Resources Association (ELRA).

Hamdy Mubarak. 2018. Build fast and accurate lemmatization for Arabic. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan, May. European Language Resources Association (ELRA).

Peter F Oliva, 2005. *Developing the curriculum (6th ed.)*. Pearson, Boston.

Sandy Soto. 2015. An analysis of curriculum development. *Theory and Practice in Language Studies*, 5:1129, 06.

R. W. Tyler. 1950. *Basic Principles of Curriculum and Instruction*. University of Chicago Press.

Wajdi Zaghouani, Behrang Mohit, Nizar Habash, Ossama Obeid, Nadi Tomeh, Alla Rozovskaya, Noura Farra, Sarah Alkuhlani, and Kemal Oflazer. 2014. Large scale arabic error annotation: Guidelines and framework. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, Reykjavik, Iceland, may. European Language Resources Association (ELRA).

Nasser Zalmout, Hind Saddiki, and Nizar Habash. 2016. Analysis of foreign language teaching methods: An automatic readability approach. In *Proceedings of the 3rd Workshop on Natural Language Processing Techniques for Educational Applications (NLPTEA2016)*, pages 122–130, Osaka, Japan, December. The COLING 2016 Organizing Committee.