

Article

Ensuring Inclusion and Diversity in Research and Research Output: A Case for a Language-Sensitive NLP Crowdsourcing Platform

Dimah Alahmadi ^{1,*} , Amal Babour ¹ , Kawther Saeedi ¹  and Anna Visvizi ² 

¹ Department of Information Systems, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah 21589, Saudi Arabia; ababor@kau.edu.sa (A.B.); ksaeedi@kau.edu.sa (K.S.)

² Effat College of Business, Effat University, Jeddah 34689, Saudi Arabia; avisvizi@gmail.com

* Correspondence: dalahmadi@kau.edu.sa; Tel.: +966-12-695-2000 (ext. 26157)

Received: 17 August 2020; Accepted: 4 September 2020; Published: 8 September 2020



Abstract: In the context of the debate on the need to place citizens at the center of the technological revolution, this paper makes a case for a natural language processing (NLP) crowdsourcing platform that ensures inclusion and diversity, thus making the research outcome relevant and applicable across issues and domains. This paper also makes the case that by enabling participation for a wide variety of stakeholders, this NLP crowdsourcing platform might ultimately prove useful in the decision- and policy-making processes at city, community, and country levels. Against the backdrop of the debates on artificial intelligence (AI) and NLP research, and considering substantial differentiation specific to the Arab language, this paper introduces and evaluates an Arab language-sensitive NLP crowdsourcing platform. The value of the platform and its accuracy are measured via the System Usability Scale (SUS), where it scores 72.5, i.e., above the accepted usability average. These findings are crucial for NLP research and the research community in general. They are equally promising in view of the practical application of the research findings.

Keywords: human-centric technology; inclusion; decision-making; natural language processing (NLP); NLP crowdsourcing platforms

1. Introduction

Successive advances in information and communication technology (ICT) have allowed unprecedented progress in the domain of crowdsourcing platforms. The basic idea behind crowdsourcing platforms is that complex tasks can be disintegrated into a great number of smaller pieces and delegated to a great number of individuals to address them, thus enabling a collective, but not necessarily conscious and optimally coordinated effort of complex problem solving. Critical in the debate on crowdsourcing platforms is the notion of task delegation, i.e., human intelligence tasks (HITs). HITs are a form of micro-work [1]. Indeed, individuals performing such tasks are usually paid or remunerated for their input in alternative ways. Importantly, the great number of individuals dealing with the small tasks assigned to them use their cognitive skills to solve problems. The case of crowdsourcing platforms suggests, however, that micro-work is a necessary structural component of contemporary artificial intelligence production processes [1].

Research on natural language processing (NLP), artificial intelligence (AI), and related applications and their uses, is continuing to proliferate [2,3]. However, if the development of NLP applications is to be maintained, language differences and culturally determined language specificity need to be brought into the analysis and application development process. This means that custom-made approaches and

solutions to NLP are required to support different languages, such as Arabic, Urdu, and Persian [4]. NLP crowdsourcing platforms attest to that. Literature suggests that considerable progress has been attained in this regard, especially in relation to English and Chinese [5,6]. As a result, relatively efficient and accurate crowdsourcing platforms have been created that, at this stage, support diverse forms of research requiring data collection [7–10]. This includes academic research, market and consumer satisfaction surveys, etc. Thanks to crowdsourcing platforms, the tasks that these activities require can be performed quickly and cost-effectively, while also allowing the collection of good quality data.

With regards to NLP, crowdsourcing platforms prove useful in a variety of research projects that require text summarization, machine translation, and speech recognition. To this end, the following crowdsourcing platforms have been commonly used: Amazon's Mechanical Turk [7], CrowdFlower [8], Lionbridge [9], Prolific Academic, and others [1,10]. Although crowdsourcing platforms have a great potential for research, each of these platforms has certain limitations. For instance, Amazon Mechanical Turk is difficult to employ outside of the United States. Sometimes, the complexity of specific languages limits the pool of participants involved in a crowdsourcing activity. This would be the case of the German ClickWorker crowdsourcing platform [11], or the Chinese Zhubajie/Witmart crowdsourcing platform used as a benchmark for research in China [12].

Some progress has been attained with regards to NLP research oriented on the Arabic language [13–15]. Nevertheless, even though a few researchers sought to develop their own crowdsourcing platforms, the tendency was to employ the existing platforms, such as Amazon Mechanical Turk and CrowdFlower. The objective of this paper is to support NLP Arabic-based research. To this end, a new crowdsourcing platform, named Tashkeel, is proposed and elaborated. The name of the platform is inspired by the eight main diacritics that transform a word written with an Arabic letter into a vast range of forms and meaning. The power of those eight marks, including Fathah, Kasrah, Dhammah, Sukun, Shaddah, Tanwin (Fath), Tanwin (Kasr), and Tanwin (Dham), in enriching the Arabic language inspired us to give the platform this name, in order to have the same impact on NLP Arabic research. The argument is structured as follows. First, a review of the crowdsourcing platforms' scene employed for studies related to the Arabic language is presented. In the next step, Tashkeel's details are elaborated and an evaluation of Tashkeel is provided. It is argued that Tashkeel should be seen as a step toward the creation of comprehensive, language-sensitive crowdsourcing platforms useful and usable not only primarily for the purposes of research, but also toward the development of open platforms in the future for innovation generation, opinion aggregation, etc., in the context of smart cities, smart communities, and e-governments.

This paper is dedicated to Arabic-based NLP projects and research. Accordingly, in this paper, we develop a general-purpose Arabic crowdsourcing platform to provide opportunities to access the expertise of all types of Arabic speakers who are proficient users of different dialects and possess the necessary skills. To the best of our knowledge, this is the first general Arabic crowdsourcing platform for NLP research. As such, the added-value of this paper consists of four items: (i) Investigating the literature of Arabic NLP research that uses crowdsourcing platforms; (ii) identification of the need to use popular platforms for Arabic NLP research; (iii) conceptualization of the design for a general Arabic language platform to support Arabic NLP research; and (iv) development and evaluation of an Arabic language-sensitive crowdsourcing platform, named Tashkeel. This platform includes the interface and the requirements to empower Arabic research by indicating specific related skills, rewards, and ratings. The rest of the paper is structured as follows: Section 2 illustrates the challenges facing Arabic research and the limitations of the available frameworks; Section 3 provides an overview of the related work and the importance of Arabic crowdsourcing platforms; Section 4 introduces the Tashkeel platform, followed by a feature comparison with other well-known platforms in Section 5; In Section 6, an evaluation of the platform via a case study is presented; and Section 7 presents the conclusion and future work.

2. Arabic Language Research Challenges

2.1. Language Complexity

In terms of people using the language on a daily basis, the Arabic language is the 6th most frequently used language in the world. Substantial diversification of the language exists. Modern Standard Arabic (MSA) and Dialectal Arabic (DA) language represent the two types of prevalent Arabic language in use. MSA is the official language used in written materials and formal communications, while DA is used by Arabs for spoken and informal communication. Five major dialects can be distinguished in the Arabic language. These include the following dialects: The Gulf, the Levantine, the Egyptian, the North African, and the Iraqi dialects [14]. In the context of NLP research, the specific differences that differentiate the MSA and DA are important to note and address. DA is distinguishable from MSA lexically, phonetically, morphologically, and orthographically. For example, lexically, *عشان* ('because') is a DA word created by combining two MSA words, which are *على* and *شأن*. Phonetically, the word *لك* ('for you') in MSA could be pronounced in Gulf Dialectal Arabic (Gulf DA) as/lis/in a feminine case and/lits/in a masculine case. Morphologically, *+ع/a+/in* DA is an abbreviated form of the preposition *على/ala/in* MSA, which can be written in DA as *عاليسار* and in MSA as *اليسار على* ('on the left'). In terms of orthography, *مئة* in MSA is written as *مئة* in DA [16]. Moreover, MSA may be difficult to understand to many people due to its old, printed manuscripts, which may hinder comprehension on account of the aging of manuscripts, unclear lettering, and color fading [17]. Irrespective of these challenges, considerable Arabic language NLP research exists. To this end, researchers have either used the already existing crowdsourcing platforms to annotate specific tasks [18–20], or built their own crowdsourcing systems. The latter were customized to satisfy the specific research objectives that respective researchers addressed [14,21,22].

2.2. Current Platform Limitations

Several limitations can be revealed in the current crowdsourcing platforms in relation to Arabic language. First, the popular crowdsourcing platforms, e.g., Amazon's Mechanical Turk or the German ClickWorker, have been developed in a specific linguistic context to draw from and to serve specific audiences. In other words, they basically communicate with specific foreign audiences in third countries, such as the US, the UK, or Germany. In the context of Arabic language NLP, it implies that research cannot access the opportunity to communicate with a wide range of Arab-speaking audiences. Second, the platforms do not provide any clarity on the level or type of spoken Arabic. In cases like this, this problem is bypassed by establishing the values 'manually', i.e., by researchers and users. It suggests, however, that the greatest promise of a crowdsourcing platform, i.e., the capacity to reach a great number of users and high rate of accomplished HITs, is thus compromised. Third, the undetermined question of language proficiency causes problems on the supply- and demand-side of the tasks of a given crowdsourcing platform; that is, on the one hand, the accuracy level of the tasks, as individual HITs, can be limited. As a result, on the demand-side, the problem is that the outcomes of a given crowdsourcing project cannot be justified with the expected and required degree of accurateness. Even in the case of crowdsourcing platforms developed for Arabic audiences, the problem is that they usually tend to target specific country dialects or specific goals, so the results cannot be generalized for broader research in the context of Arabic language. Irrespective of these challenges, substantial studies in the field of Arabic language NLP using crowdsourcing have been reported. The following section sheds light on these works.

3. Related Work

Considerable work has been done in the context of crowdsourcing and Arabic language in recent years. Some of these studies have utilized existing crowdsourcing platforms, such as Amazon

Mechanical Turk and CrowdFlower, while others have used self-created platforms. The following paragraphs elaborate on these.

With reference to annotating Arabic dialect in a sentence, Zaidan and Callison Burch [19] relied on Amazon Mechanical Turk crowdsourcing to give labels to a randomly selected set of about 110,000 Arabic sentences. Those sentences are chosen from over three million sentences of the Arabic On-line Commentary dataset (AOC). AOC is a combination of three newspapers, including *Al-Ghad* from Jordan, *Al-Riyadh* from Saudi Arabia, and *Al-Youm Al-Sabe'* from Egypt. The annotators were given short and simple instructions to label the sentences, wherein each label had to include details about the level and type of dialect in each sentence. The level indicated the extent of Arabic dialect in the sentence in accordance with available options. These included no dialect for the pure MSA sentence, a small amount of dialect, a mixed amount of dialect and MSA, an extensive amount of dialect, and a non-Arabic sentence. The type referred to which Arabic dialect was manifested in the sentence, whether Gulf, Egyptian, Levantine, Iraqi, or Maghrebi. The authors randomly grouped the sentences into sets of ten sentences. Each group was presented to each annotator on a single screen. Two more MSA sentences were randomly selected and added to each screen as control sentences to test the annotators' accuracy. These two sentences had to be labeled as MSA. Any annotator who mislabeled them was considered a spammer and their work was rejected. Three different annotators were allocated to label the sentences on each screen.

Alsarsour and his associates [15] created a large annotated multi-dialect dataset of about 25 thousand Arabic tweets called Dialectal Arabic Tweets (DART). For creating DART, they tapped into a list of Arabic words collected by [23] and a list of Arabic phrases from the Mo3jam website using Twitter streaming Application Programming Interface (API). They used the CrowdFlower crowdsourcing platform to annotate each tweet, wherein the annotation label had to indicate the type Arabic dialect (Egyptian, Maghrebi, Levantine, Gulf, or Iraqi) used. To test the quality, the authors asked five native speakers, including one for each dialect group, to label about 300 to 400 tweets. They used this set as a test source to randomly select 10 tweets to test the annotators before they commenced the job. The annotators needed to get a minimum of 90% to pass the test. If the annotators achieved a score of below 90%, they were excluded.

Another team of authors [24] created a corpus concentrating on the Levantine dialect used in Levantine countries. They used Twitter to create a Levantine dataset including 4000 tweets. They looked for the following information: (i) The overall sentiment of the tweet on a 5-point scale; (ii) the target to which the sentiment was expressed; (iii) how the sentiment was expressed; and (iv) the topic of the tweet. The researchers used the CrowdFlower platform. They provided the annotators with guided instructions to determine the overall sentiment and to select the target of this sentiment within the tweet. In addition, the annotators had to identify whether the sentiment was obtained explicitly or implicitly. The annotators also had to explain the topic included from predefined topics.

The four topics were presented as follows: Politics, religions, sports, and personal. If the topic was not included amongst the four topics given to them, the annotators had to specify their own opinion. Before the platform was used for the entire task, the clarity of the instructions was tested by applying a pilot task. A number of different annotators (from 5 to 9) were allocated for each tweet. The annotators' accuracy on the platform was tested by comparing it against 181 tweets annotated by the authors as the gold standard. Only annotators with an accuracy higher than 75% were accepted for the large-scale task, whereas the others were rejected.

For annotating targets of opinion in Arabic, Fara et al. [25] utilized the Amazon Mechanical Turk platform to annotate Arabic comments about Aljazeera newspaper articles chosen from the Qatar Arabic Language Bank (QALB). They chose a randomly selected sample of 1177 comments on articles about politics, sports, and culture topics for annotation. The process was divided into two different stages. Each had a series of tasks. The task instructions were written in Arabic. In the first stage, a comment was given to three annotators. They were asked to identify whether the entities in a comment represented nouns referring to people, places, things, or ideas. Their answers, if overlapping,

were then used to refer to the entity in the comment. In the second stage, a comment with a single entity was given to five annotators. They were asked to identify the opinion about the entity in the comment in terms of whether it was positive, negative, or neutral.

Furthermore, by employing social media, [18] created a corpus of 76,619 Arabic tweets mentioning forms of violence and abuse. They prepared a list of 237 words in Arabic denoting violence and tracked them using Twitter streaming API to collect the corpus. They used the CrowdFlower platform to assign annotations to each tweet in the corpus. Each tweet had to be annotated by at least five annotators. The annotation option referred to one of the seven types of violence, including human rights laws (HRA), political opinion, accidents, crime, conflict, crises, and violence, amongst others. They randomly selected a set of 206 tweets from the corpus and had these annotated manually by experts. This set was used as a test source to test the annotators' performance before they started the job. The annotators needed to get at least 70% to continue working on the task. From the corpus, they annotated 20,151 tweets.

For annotating media content, Al-Muzaini and Al-Yahya [26] created part of the Flickr and MS COCO caption dataset in an Arabic version as the first public Arabic image caption corpus. The corpus consists of 3427 images from both Flickr and MS COCO. The MS COCO dataset consists of 330,000 images with 2.5 M captions. The authors selected 1166 images with 5358 captions from the dataset. They utilized the CrowdFlower crowdsourcing platform as a human translator to provide Arabic captions for part of their database. Flickr8K consists of 8000 images. The authors selected 2261 images from the Flickr8K dataset. They had 150 images from Flickr with a total of 750 Arabic captions translated by professional English-Arabic translators. They translated the rest of the images by using the Google translator and then checked the translation provided by Arabic native speakers.

On the other hand, a game-based study [22] developed the Kalemah system to digitize scanned Arabic documents. A challenging game was presented to the volunteers to encourage them to type the words swiftly and correctly. Moreover, volunteers had the ability to play with friends from social media who were invited to join the game. These words were extracted from scanned Arabic documents which required transformation into a digital format. They explained how micro-tasks could be achieved easily from crowdsourcing, especially if the platform introduced the task in a game context.

In another study [21], the authors explored altruistic crowdsourcing, which is a volunteer-based crowdsourcing platform, to validate their approach named Kalam'DZ. They chose this type of crowdsourcing to harness community interest in a topic, rather than to draw people in to take up the task in pursuit of payment. They built a CrowdCrafting project that included two main tasks, i.e., Task Presenter and Task Creator. Workers who were craftsmen and employers were recruited. They created a form with target speech audio and buttons for responses about Algerian dialects, along with an Algerian map to help the craftsmen. Their study targeted only Algerian users and the written Arabic form. They showed how altruistic workers could achieve good results given that 81% of workers accurately matched the gold standard annotated by the experts.

This section has demonstrated several Arabic language research studies conducted with the help of crowdsourcing platforms. Our detailed investigation of the recent works proposed in this domain motivated the research team to propose specialized crowdsourcing for Arabic NLP research after concluding the following points.

There is a growing interest in Arabic NLP research; however, these platforms do not target Arabic audiences, which can lead to limited crowd participation. There is no platform that attracts Arabic speaking audiences using their language, and a single channel needs to be developed, which then will empower Arabic language research and unite efforts.

Although we have elaborated on a considerable body of Arabic research using the available platforms, we have also pointed out its limitations. At this point, we suggest that it is difficult to expand research objectives that highlight the unique Arabic language features and complexity.

A qualified Arabic audience with limited English language skills may have difficulties accessing the available platform and understanding the micro-task instructions.

The translation of the available interface platforms is not sufficient; however, the localization and addressing the complexity and diversity of the language are highly required.

4. Conceptualizing the Tashkeel Platform

This paper proposes the Tashkeel platform [27] as a crowdsourcing platform for Arabic language NLP research. This platform intends to support Arabic NLP language research. The platform addresses unique features and different dialects of the Arabic language, which may require specific qualifications. In addition to the features found in popular crowdsourcing platforms, Tashkeel offers an Arabic user interface, as well as new options that address the complexity of this rich language. For example, the project owner can specify the skillset needed for a project by choosing a specific dialect or qualification. This section illustrates the main aspects of the development of the Tashkeel platform.

To attain the specific research objectives thus defined, the paper employs a mixed method approach, bringing together qualitative research [28] and applied research. To establish the context and build the conceptual frame, in which the case of Tashkeel is examined, content analysis and a case study method are employed [29]. The critical evaluation of the findings thus attained is supported by the nested analysis method [30].

4.1. Tashkeel Platform Design

A use case diagram is employed to demonstrate system behavior in terms of actions that the system performs in collaboration with one or more users. The actions are illustrated in the use case, which provides observable and valuable results to system users. Figure 1 depicts Tashkeel’s core function and main users in a use case diagram. The system has two main actors, namely, the project owner and the crowd. The project owner is a researcher who wants the crowd to help in preparing or creating the dataset for a specific study. The project owner is responsible for creating the project, identifying the skills required, and most importantly creating the micro-tasks for the crowd participants. After the micro-task submission, the project owner assesses the quality of the submitted micro-tasks themselves and may accept or reject the participation, as well as rate the submitted work. The second actor is the crowd member who undertakes the micro-tasks. A crowd member is required to set up an account and specify their Arabic skill level. The system places an emphasis on three main Arabic skills, as follows:

1. The level of Arabic language, with at least one of the following characteristics or qualifications: Arabic speaker, mother language, bachelor’s degree in Arabic, or graduate studies in Arabic;
2. The dialect languages include Arabic dialects from all Arabic countries;
3. The adapted skills in Arabic include skills such as Arabic calligraphy, grammar, listening, public speaking, story-writing, and article-writing.

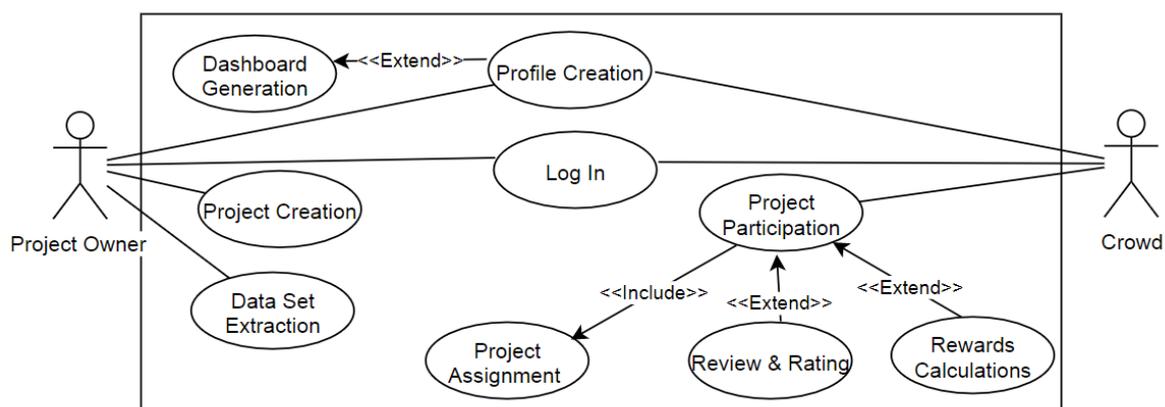


Figure 1. Tashkeel’s core function in a use case diagram.

In addition to the actors, the use case diagram above illustrates the key functions of the Tashkeel platform, such as project creation; project participation; and rewards calculation, review, and rating. The system workflow is useful for clarifying the series of necessary activities and the sequence amongst these for completing a process. The Business Process Model and Notation (BPMN) diagram is used to model a high-level workflow of Tashkeel’s main business process. The diagram illustrates the two main actors, namely, the project owner and the crowd interacting with the Tashkeel platform. The process starts when a project owner creates a project. This activity allows the project owner to configure micro-task requirements and a submission workflow. It includes setting micro-tasks, skills required, rewards, project time windows, participation permissions, and task submission. Once the project is open, the Tashkeel platform displays the list of projects and micro-tasks to the crowd. To ensure that all of the micro-tasks have the same hiring opportunity, a round assignment logic is implemented. Once the crowd adds participation, the data is stored in the project dataset, the reward/payment is calculated, and the invoice is issued. Figure 2 demonstrates Tashkeel’s main workflow using BPMN.

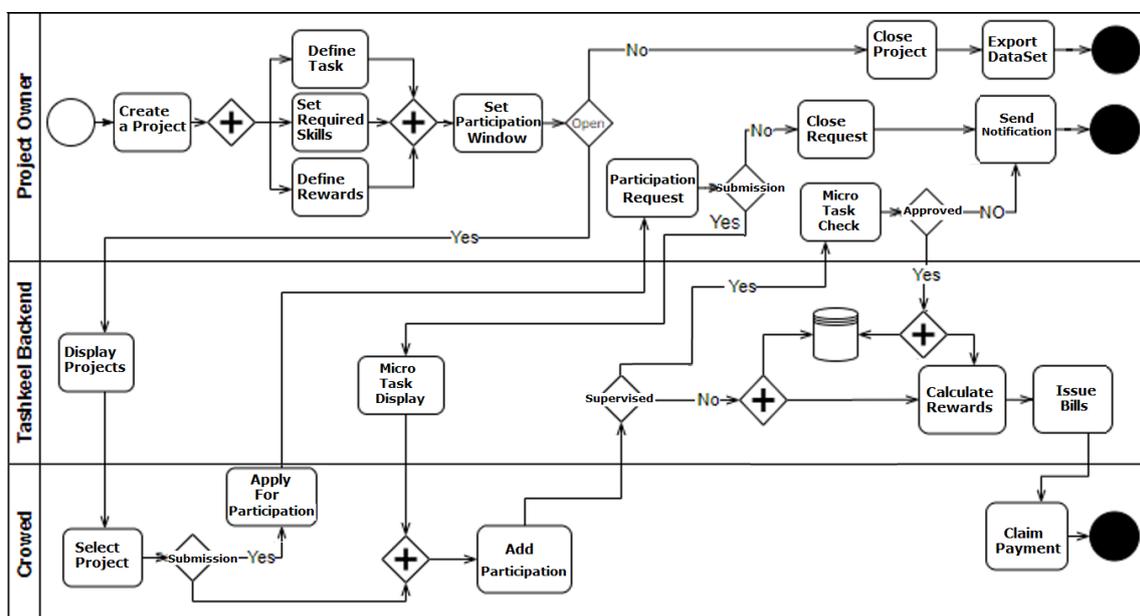


Figure 2. Tashkeel’s main workflow in a Business Process Model and Notation (BPMN) diagram.

4.2. Applying Tashkeel

Tashkeel is a web-based system implemented using Asp.net, HTML, CSS, C#, JavaScript, and MSSQL. The Arabic language is employed for the user interface of Tashkeel’s web pages. An Agile development approach governs the development process in order to speed up the development time and get faster feedback. In alignment with agile principles, each function developed is tested and reviewed by target users. This implicitly refers to unit testing, integration testing, and user acceptance testing for each function developed. The remaining part of this section demonstrates important screenshots demonstrating the functioning of the Tashkeel platform.

The project owner is the main actor of the system, wherein project creation is the main activity in the Tashkeel platform. The project owner needs to specify the type of NLP projects using seven available types:

1. Image classification/الصور تصنيف: Classifies images based on specific questions;
2. Conversion from audio to text/نص إلى صوت من تحويل: Converts speech in an audio clip into written text;
3. Conversion from text image to editable text/نص إلى صورة من تحويل: Converts text in an image into written text;

4. Translation from dialect Arabic to Modern Standard Arabic/فصحى عربي الى عربي من ترجمة: Translates text written in dialect Arabic into Modern Standard Arabic;
5. Translation from any language to Arabic/عربي الى لغة اي من ترجمة: Translates text written in any language into Arabic;
6. Text classification/النصوص تصنيف: Classifies text based on specific questions;
7. Sound classification/الأصوات تصنيف: Classifies sound based on specific questions.

Figures 3–6 illustrate screenshots for project type specification. Please note that the screenshots depict figures described in Arabic characters, precisely because we have developed an Arabic language crowdsourcing platform. We are aware of the challenge that non-Arab readers will encounter at this stage. To bypass this challenge, under each of these figures, we provide a brief description of the content and purpose of the respective figure.

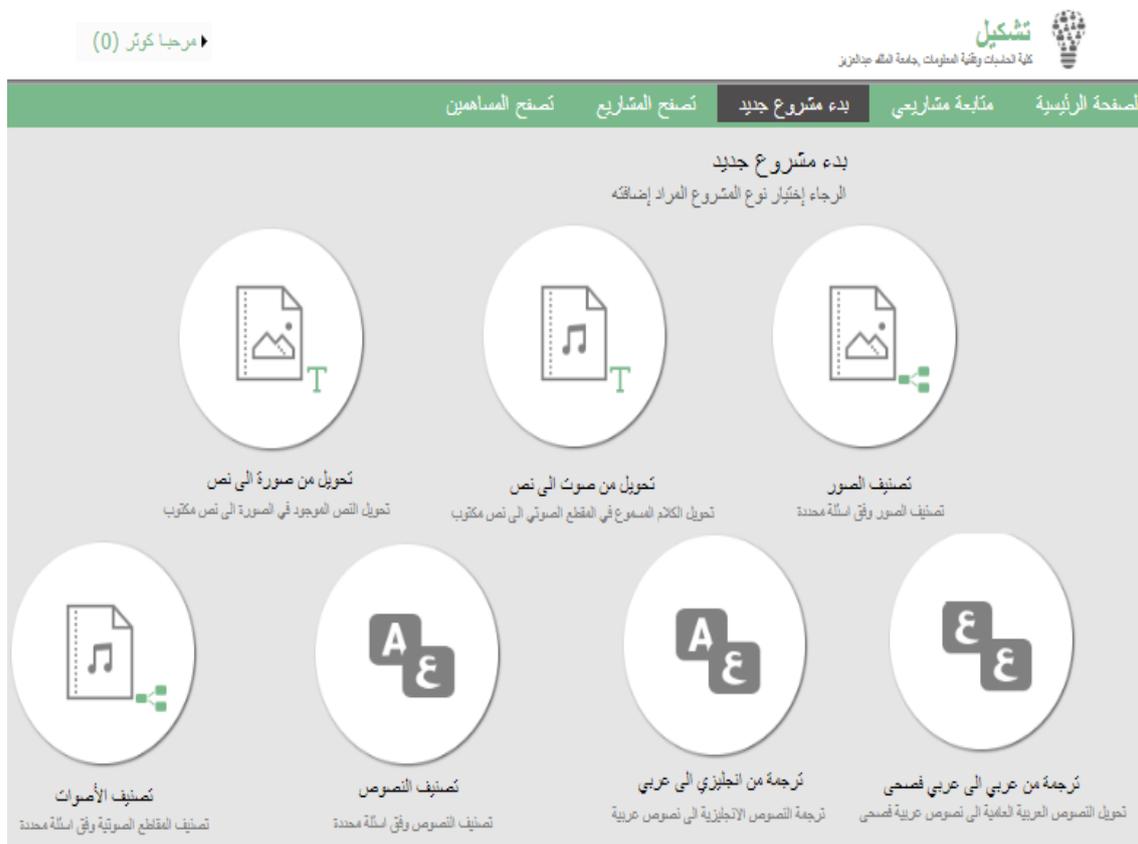


Figure 3. Tashkeel natural language processing (NLP) project types (this figure shows seven types of NLP projects, as described in point 1 to 7 in the previous paragraph).

The project owner provides a description and instructions to clarify the tasks, skills required, and rewards, as shown in Figure 4. The project owner can click to view the participation details, review and rate the participation, and perform other actions related to the work conducted. In addition, a search and filter using different fields, including the Arabic level, Arabic dialect, and project type, to view the crowd profiles available, as shown in Figure 5, is available. In addition, the project owner dashboard where they can navigate the open and closed projects, task worker requests, submissions, and work pending approval is shown in Figure 6.

بدء مشروع جديد

إنشاء مشروع جديد يطلب من المساهم ترجمة النصوص العربية العامية الى عربية فصحي

عنوان المشروع *

وصف المشروع * اكتب نبذة بسيطة عنك مع شرح المطلوب بشكل مختصر

مستوى اللغة العربية المطلوب * متحدثي اللغة

اللهجات المطلوبة *

جميع اللهجات العربية

سعودية

إماراتية

بحرينية

كويتية

لهجة أخرى:

المهارات المطلوبة *

الخط العربي

النحو والصرف

الإستماع

الإلقاء

كتابة قصة

مهارات أخرى:

عدد النقاط * يوم **الحد الأقصى لإنجاز المهمة ***

Figure 4. Project setting.

عرض جميع المساهمين

تصفح جميع المساهمين في المنصة. لرؤية المشاريع الخاصة بالمساهم وملفه الشخصي الرجاء الضغط على ايقونة الملف امام المساهم.

الرقم التعريفي	اسم المساهم	الدولة	التقييم	تاريخ الاضمام	مستوى اللغة	المهارات اللغوية	اللهجات المتقنة	الملف الشخصي
4	Amal	السعودية	☆☆☆☆☆ 4	Dec 1 2019	متحدثي اللغة	النحو والصرف	سعودية	
6	بنان	السعودية	☆☆☆☆☆ 0	Dec 3 2019	اللغة الام .. مستوى عادي	الخط العربي, النحو والصرف, الإستماع	سعودية	
21	حنين	السعودية	☆☆☆☆☆ 4	Dec 11 2019	اللغة الام .. مستوى عادي	الإستماع, الإلقاء	سعودية	
22	منار	السعودية	☆☆☆☆☆ 4	Dec 11 2019	متحدثي اللغة	الإستماع, الإلقاء	سعودية	
23	فاطمة	السعودية	☆☆☆☆☆ 4	Dec 11 2019	متحدثي اللغة	جميع اللهجات العربية	سعودية	

Figure 5. Contributor skills filter result.

المشاريع الحالية

متابعة جميع المشاريع الحالية المفتوح فيها إستقبال طلبات العمل. لإيقاف إستقبال الطلبات الرجاء الضغط على ايقونة الإيقاف امام المشروع المراد إيقافه

رقم المشروع	اسم المشروع	عدد النقاط	المدة	تاريخ الإنشاء	عدد المساهمين	عدد الطلبات	عرض الطلبات	تفاصيل المشروع	عرض النتائج	إيقاف الطلبات	تعديل المشروع
220056	ترجمة مصطلحات متعلقة بالحاسب والتقنية من اللغة الانجليزية الى اللغة العربية 1	5	1	Dec 10 2019	5	0	عرض الطلبات	تفاصيل المشروع	عرض النتائج	إيقاف الطلبات	تعديل المشروع
220057	ترجمة مصطلحات متعلقة بالحاسب والتقنية من اللغة الانجليزية الى اللغة العربية 2	5	1	Dec 10 2019	5	0	عرض الطلبات	تفاصيل المشروع	عرض النتائج	إيقاف الطلبات	تعديل المشروع
220058	ترجمة مصطلحات متعلقة بالحاسب والتقنية من اللغة الانجليزية الى اللغة العربية 3	5	1	Dec 10 2019	5	0	عرض الطلبات	تفاصيل المشروع	عرض النتائج	إيقاف الطلبات	تعديل المشروع
220059	ترجمة مصطلحات متعلقة بالحاسب والتقنية من اللغة الانجليزية الى اللغة العربية 4	5	1	Dec 10 2019	3	0	عرض الطلبات	تفاصيل المشروع	عرض النتائج	إيقاف الطلبات	تعديل المشروع
220060	ترجمة مصطلحات متعلقة بالحاسب والتقنية من اللغة الانجليزية الى اللغة العربية 5	5	1	Dec 10 2019	1	0	عرض الطلبات	تفاصيل المشروع	عرض النتائج	إيقاف الطلبات	تعديل المشروع

Figure 6. Project dashboard.

Figure 4 depicts the settings and criteria required in a new project. The project owner will fill in fields such as the project title (عنوان المشروع), a description of the project (وصف المشروع), and the required level of qualification in Arabic language (مستوى اللغة العربية المطلوبه). These may include at least one of the following characteristics or qualifications: Arabic speaker (متحدثي اللغة), mother language (الغه الأم), bachelor's degree in Arabic (بكالوريوس لغه عربيه), or graduate studies in Arabic (شهادة دراسات عليا في اللعربية). Next, the owner will choose the specific Arabic dialects (اللهجة المطلوبه) from all Arabic countries. The project owner can also include skills, such as Arabic calligraphy (الخط العربي), grammar (النحو), listening (الاستماع), public speaking (الالقاء), and story-writing (الحد الاقصى لانجاز المهمه) in the project setting, the micro task duration is specified in (الحد الاقصى لانجاز المهمه).

Figure 5 shows all contributors listed and active in the platform. The owner can navigate through their profiles (الملف الشخصي). In this way, the project owner can learn the contributors' names (اسم المساهم), country (الدولة), ratings (التقييم), date of joining the platform (تاريخ الانضمام), Arabic language qualifications (مستوى اللغة), skills (المهارات اللغويه), and spoken dialect (اللهجة المتعلقه).

Figure 6 depicts the screen featuring the dashboard that the project owner will use to follow the projects they are working on at a given time (المشاؤيع الحالية). Accordingly, it serves as an overview of the project ID (رقم المشروع), the title of the project (عنوان المشروع), the date (تاريخ الانشاء), and the number of contributors (عدد المساهمين). The owner can also navigate the requests, the project details (المشروع تفاصيل), and the project results (عرض النتائج). The owner can update the project (تعديل المشروع) or stop the requests (إيقاف الطلبات).

The project owner can receive a generated dataset exported in an Excel file as a result of each of the seven NLP project types offered. The project owner can obtain a classified image, Arabic text of converted audio, Arabic text of converted images, text of standard Arabic translated from Arabic dialect, Arabic text translated from any language, classified Arabic text, or a classified sound dataset based on the project. Figure 7 presents a sample of a generated dataset of Arabic text translated from the English language.

رقم النص	النص المراد ترجمته	النص المترجم
232	antivirus	فيروس
232	antivirus	برامج للحماية ضد الفيروسات
232	antivirus	مكافحة الفيروسات
232	antivirus	مضاد الفيروسات
232	antivirus	برنامج حماية
233	artificial intelligence	ذكاء اصطناعي
233	artificial intelligence	الذكاء الاصطناعي.
233	artificial intelligence	الذكاء الاصطناعي
233	artificial intelligence	ذكاء اصطناعي
233	artificial intelligence	الذكاء الاصطناعي
234	Bluetooth	بلوتوث
234	Bluetooth	اتصال لا سلكي بين الأجهزة
234	Bluetooth	نقل بيانات

Figure 7. Generated dataset.

Figure 7 depicts examples of the results of a project listed with the test number (رقم النص). It explains the case of translation from English into Arabic. Here specifically, we see the term 'antivirus' in text for the translation column (النص المراد ترجمته) and the five translation attempts into Arabic in the translated column (النص المترجم) (فيروس, برامج للحماية ضد الفيروسات, مكافحة الفيروسات, مضاد الفيروسات و برنامج حماية). Another important actor in the crowd (مساهم/مساهم) can view the posted projects, related micro-tasks, and their rewards. After a successful login, the contributor can view a dashboard of previous work performed, which also takes them to the participation details. In order to participate, a list of open projects with the desired skills are listed, as shown in Figure 8, and more detailed information can be accessed by clicking the project details. The project can be filtered by the skills required, project types, and other properties. The contributors from the crowd can work on the chosen micro-task if access permission is given. Figure 9 shows an example of a micro-task screen for English to Arabic translation.

The screenshot shows a web interface for project participation. The main content area lists four projects, each with the following details:

- Project 1:** ترجمة مصطلحات متعلقة بالحاسب والتقنية من اللغة الإنجليزية الى اللغة العربية 5. 20 مصطلح من مصطلحات متعلقة بالحاسب والتقنية من اللغة الإنجليزية الى اللغة العربية. 1 يوم, 5 نقطة. رقم المشروع: 220060. تاريخ الإنشاء: Dec 10 2019.
- Project 2:** ترجمة مصطلحات متعلقة بالحاسب والتقنية من اللغة الإنجليزية الى اللغة العربية 4. 20 مصطلح من مصطلحات متعلقة بالحاسب والتقنية من اللغة الإنجليزية الى اللغة العربية. 1 يوم, 5 نقطة. رقم المشروع: 220059. تاريخ الإنشاء: Dec 10 2019.
- Project 3:** ترجمة مصطلحات متعلقة بالحاسب والتقنية من اللغة الإنجليزية الى اللغة العربية 3. 20 مصطلح من مصطلحات متعلقة بالحاسب والتقنية من اللغة الإنجليزية الى اللغة العربية. 1 يوم, 5 نقطة. رقم المشروع: 220058. تاريخ الإنشاء: Dec 10 2019.
- Project 4:** ترجمة مصطلحات متعلقة بالحاسب والتقنية من اللغة الإنجليزية الى اللغة العربية 2. 20 مصطلح من مصطلحات متعلقة بالحاسب والتقنية من اللغة الإنجليزية الى اللغة العربية. 1 يوم, 5 نقطة. رقم المشروع: 220057. تاريخ الإنشاء: Dec 10 2019.

The right sidebar contains the following filters:

- بحث عن مشروع:** عدد المشاريع: 5
- نوع المشروع:**
 - تحويل من صورة الى نص
 - تحويل من صوت الى نص
 - تصنيف الصور
 - تصنيف الأصوات
 - تحويل من الإنجليزي الى عربي
 - تحويل من عربي الى عربي
 - تصنيف النصوص
- اللهجات:**
 - جميع اللهجات العربية
 - سعودية
 - إماراتية
 - بحرينية
 - كويتية
 - خليجية
- مستوى اللغة العربية:**
 - متحدثي اللغة
 - اللغة الام .. مستوى عادي
 - متخصص .. بكالوريوس
 - متخصص .. دراسات عليا

Figure 8. Open project for the participation page.

owner country, but the list does not display any of the Arab countries except Qatar. This hinders the benefit of using the platform in other Arab countries. In addition, there is no required fill-in information about the language used and required on the project specification page, as can be seen in Figure 11.

Create an Amazon Mechanical Turk Account

Contact Information

Please select existing contact information or enter new information below
* Required

Company name: * (See details)

Address: *

City: *

State, province or region:

Zip or postal code: *

Country: *
United States

Phone number: *

Your phone number may be used to verify your identity.

Tell us how you plan to use MTurk:

- Conduct surveys or market research
- Annotate data sets to train machine learning models
- Collect data sets, e.g. gather the contact info for a list of businesses
- Moderate user-generated content, e.g. comments, images, or videos
- Clean or categorize data for data analysis
- My use for MTurk isn't listed above

How often do you expect to use MTurk?

- I have a one-time need, e.g. a specific project
- I have a recurring need, e.g. processing data regularly
- I will occasionally use MTurk

By creating an account, you agree to Amazon Mechanical Turk's [Participation Agreement](#). Please see our [Privacy Notice](#).

Create Account

Figure 10. Client account in Amazon Mechanical Turk.

Select a customizable template to start a new project

[Survey](#)

- Survey Link
- Survey

Vision

- Image Classification
- Bounding Box
- Semantic Segmentation
- Instance Segmentation
- Polygon
- Keypoint
- Image Contains
- Video Classification
- Moderation of an Image
- Image Tagging
- Image Summarization

Language

- Sentiment Analysis
- Intent Detection
- Collect Utterance
- Emotion Detection
- Semantic Similarity
- Audio Transcription
- Conversation Relevance
- Document Classification
- Translation Quality
- Audio Naturalness

Other

- Data Collection
- Website Collection
- Website Classification
- Item Equality
- Search Relevance

Edit Project

1 Enter Properties 2 Design Layout 3 Preview and Finish

Project Name: Survey Link This name is not displayed to Workers.

Describe your survey to Workers

Title Answer a survey about your opinions

Describe the survey to Workers. Be as specific as possible, e.g. "answer a survey about movies", instead of "short survey", so Workers know what to expect.

Description Give us your opinion about our products

Give more detail about this survey. This gives Workers a bit more information before they decide to view your survey.

Keywords survey, demographics

Provide keywords that will help Workers search for your tasks.

Setting up your survey

Reward per response \$

This is how much a Worker will be paid for completing your survey. Consider how long it will take a Worker to complete your survey.

Number of respondents 20

How many unique Workers do you want to complete your survey?

Time allotted per Worker 1 Hours

Maximum time a Worker has to complete the survey. Be generous so that Workers are not rushed.

Survey expires in 7 Days

Maximum time your survey will be available to Workers on Mechanical Turk.

Auto-approve and pay Workers in 3 Days

This is the amount of time you have to reject a Worker's assignment after they submit the assignment.

Figure 11. Project type specification in Amazon Mechanical Turk.

Figure 12 shows an easy and quick sign up to Appen and selection of the project type. It is clear that there are no specific details that can be recorded about the language required in the project from the start. There is no indication of language constraints in Appen when creating the account.

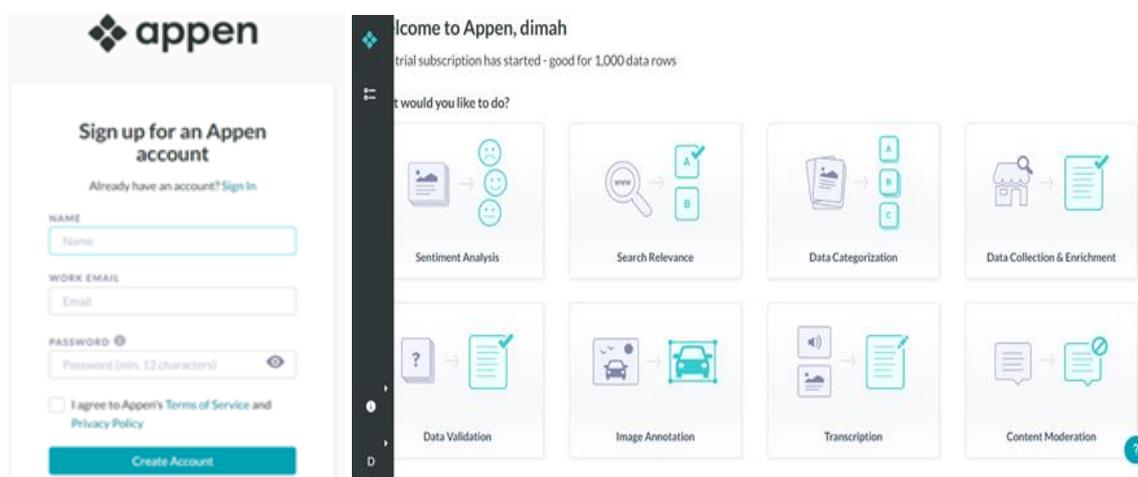


Figure 12. Client account and project in Appen.

5.2. Contributors' Features

This section illustrates how these platforms allow the project owner to set up conditions and specific features for contributors. In Amazon Mechanical Turk, the project owner can specify any qualification that workers must meet to work on the project from a list. Regarding the languages, only five languages are available in the list, not including Arabic language, as shown in Figure 13. In addition, there is no option to add more specifications or qualifications required for the project other than the listed ones.

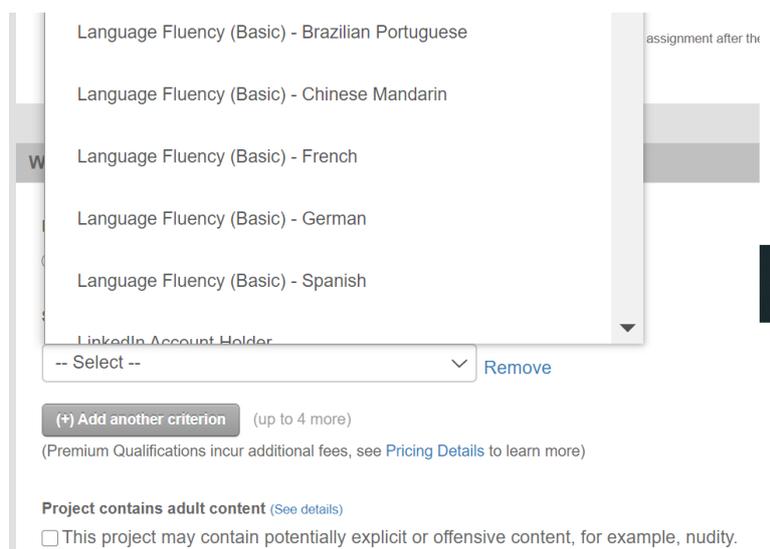


Figure 13. Specification languages in Amazon Mechanical Turk.

The Appen platform gives the project owner options to choose contributors' features. From Figure 14 regarding language, there is an option to choose the Arabic language and the country. There are very limited constraints for Arabic language and it ignores the variety of dialects, level of qualifications, and skills that most recent Arabic NLP research requires.

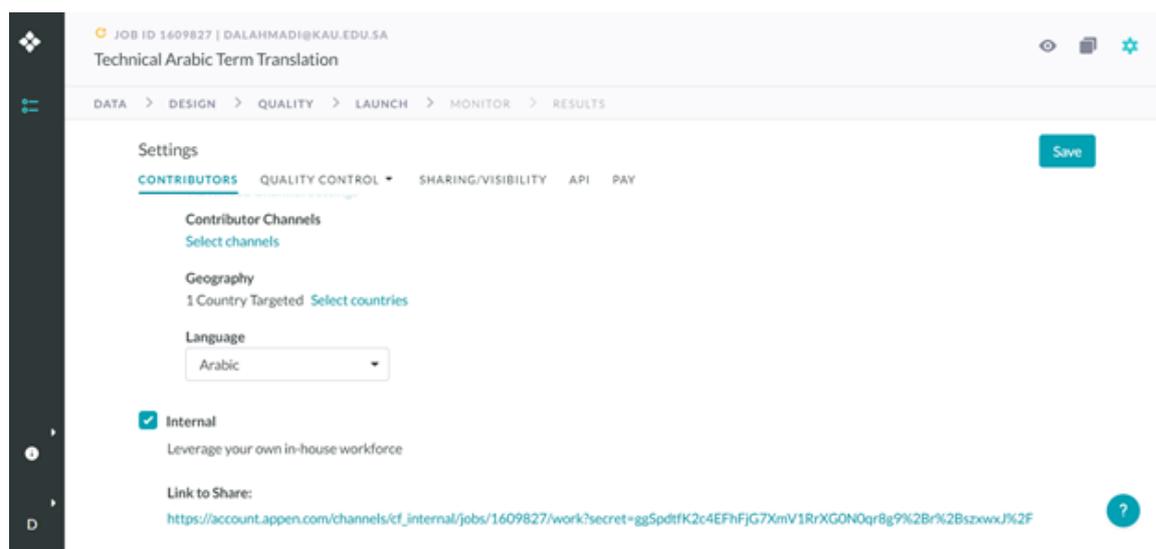


Figure 14. Limited contributors' features for Arabic language in Appen.

6. Tashkeel Evaluation: Technical Term Translation Case

For the purpose of testing Tashkeel, i.e., the Arabic language NLP crowdsourcing platform conceptualized, designed, developed, and elaborated in this paper, a technical term translation exercise was conducted at the King Abdulaziz University (KAU). The system was tested with real-world users. For this exercise, a beta version of the platform was launched using a remote server and a real case study of the Technical Term Dictionary (المعجم العربي للمصطلحات التقنية) was conducted. The Tashkeel platform supports different types of problems that require human intelligence tasks to be handled by crowdsourcing. For the purpose of evaluation, the case study conducted was titled Arabic Technical Terms Dictionary (المعجم العربي للمصطلحات التقنية) in a guided setting, in order to observe user interaction with the system followed by the performance of a survey. Two workflows were written in detailed steps: One for the project owner and the second for the crowd participant. The project owner was asked to create an account and then to create a project of translation from any language to Arabic (ترجمة من اي لغة الى عربي), followed by the uploading of an Excel sheet containing 100 technical words randomly selected from the Techopedia website [37]. In this case, the project owner was asked to define 20 words as the minimum number of words to be translated for every micro-task. Therefore, the selected 100 words were divided into five micro-tasks, each containing 20 words. Then, the project owner was asked to publish the project for a one-week period.

On the other end, the crowd contributors were asked to create accounts and access the project listed for this case study. Twenty participants from the Faculty of Computing and Information Technology (FCIT) at KAU served as contributors. Table 1 summarizes the participants' demographics.

Table 1. A summary of participants' demographics.

Age Range	+25
Specialization	
Computer science Faculty	4
Computer science Master's students	20
Technology background	
Windows pc	20

They followed the instructions given to navigate the platform and undertake the micro-tasks. After conducting the micro-tasks, a survey containing the ten questions found on the SUS questionnaire [31,32] was administered, as shown in Table 2.

Table 2. System Usability Scale (SUS) questions used in the survey [31].

1.	I think that I would like to use this system frequently.
2.	I found the system unnecessarily complex.
3.	I thought the system was easy to use.
4.	I think that I would need the support of a technical person to be able to use this system.
5.	I found the various functions in this system were well integrated.
6.	I thought there was too much inconsistency in this system.
7.	I would imagine that most people would learn to use this system very quickly.
8.	I found the system very cumbersome to use.
9.	I felt very confident using the system.
10.	I needed to learn a lot of things before I could get going with this system.

Source: Authors.

Based on the responses collected from the 20 participants, the SUS score obtained was 72.5. This means that the Tashkeel platform was found to have a higher usability than indicated by the average score attained by other platforms. Conducting the SUS in this case study was useful for confirming the importance of the Tashkeel platform in fulfilling the requirements of NLP Arabic research. This has been proven through feedback from researchers who used the beta version of the platform to create a dataset for Arabic sign language translation. Furthermore, the case study highlighted some areas of improvement for enhancing the usability of the platform.

7. Conclusions and Future Work

The objective of this paper was to address the challenge of a lack of a comprehensive Arabic language-sensitive NLP crowdsourcing platforms. The review of existing approaches to crowdsourcing platforms, including the internationally popular crowdsourcing platforms, or to individually designed platforms, highlighted that there was a need to build an Arabic language-sensitive NLP crowdsourcing platform. Accordingly, the platform, named Tashkeel, was built and evaluated via SUS. The score was 72.5, i.e., well above the accepted usability average. Against this backdrop, a case was made that, if further developed and popularized, this crowdsourcing platform might encourage research requiring this tool of research to connect researchers and users in the linguistic domain of the Arabic language. Indeed, the crowdsourcing platform developed in this paper allows Arabic speakers with different levels of proficiency and qualifications to be identified and invited to participate in a research project from its inception. Any Arabic dialects can be chosen and dedicated to the project. In this regard, on the one hand, this crowdsourcing platform provides NLP researchers with new spaces to interact with a highly diversified, and thus, 'representative' 'crowd', and to perform specific HITs. Hence, research utilizing this NLP crowdsourcing platform acquires validity, relevance, and usability. This feature of the NLP crowdsourcing platform introduced in this paper defines this platform's value-added in the field of data-driven decision-making and policy processes in general. Certainly, the road from this NLP crowdsourcing platform to the decision-making process is a long one; however, without the platform, masses of important pieces of information might be ignored. We are aware that further work needs to be done to address the issue of quality control as regards the specific HITs, e.g., by means of an algorithm that measures the quality by testing the accuracy of the participants' answers. Nevertheless, we also want to stress that there is an urgent need to pursue, in parallel, research on the value-added of crowdsourcing platforms for the decision-making process. These two issues shall be addressed in our future research.

Author Contributions: Conceptualization, D.A., A.B., K.S., and A.V.; methodology, D.A., A.B., and K.S.; software D.A., A.B., and K.S.; validation, D.A., A.B., and K.S.; writing—original draft preparation, D.A., A.B., and K.S.; writing—review and editing, D.A., A.B., K.S., and A.V. All authors have read and agreed to the published version of the manuscript.

Funding: This project was funded by the Deanship of Scientific Research (DSR), King Abdulaziz University, Jeddah, Saudi Arabia, under grant G: 1473-612-1440 and grant No. 62467. The authors D.A., A.B., and K.S. therefore gratefully acknowledged the DSR technical and financial support.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Tubaro, P.; Casilli, A.A.; Coville, M. The trainer, the verifier, the imitator: Three ways in which human platform workers support artificial intelligence. *Big Data Soc.* **2020**, *7*, 2053951720919776. [CrossRef]
2. Visvizi, A.; Jussila, J.; Lytras, M.D.; Ijäs, M. Tweeting and mining OECD-related microcontent in the post-truth era: A cloud-based app. *Comput. Hum. Behav.* **2020**, *107*, 105958. [CrossRef]
3. Mora-Cantalops, M.; Sánchez-Alonso, S.; Visvizi, A. The influence of external political events on social networks: The case of the Brexit Twitter Network. *J. Ambient. Intell. Hum. Comput.* **2019**, 1–13. [CrossRef]
4. Dashtipour, K.; Gogate, M.; Li, J.; Jiang, F.; Kong, B.; Hussain, A. A hybrid Persian sentiment analysis framework: Integrating dependency grammar based rules and deep neural networks. *Neurocomputing* **2020**, *380*, 1–10. [CrossRef]
5. Miller, T.; Turković, M. Towards the automatic detection and identification of English puns. *Eur. J. Humour Res.* **2016**, *4*, 59–75. [CrossRef]
6. Chen, L.; Song, L.; Shao, Y.; Li, D.; Ding, K. Using natural language processing to extract clinically useful information from Chinese electronic medical records. *Int. J. Med. Inf.* **2019**, *124*, 6–12. [CrossRef] [PubMed]
7. Crowston, K. Amazon mechanical turk: A research tool for organizations and information systems scholars. In *Shaping the Future of ICT Research. Methods and Approaches*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 210–221.
8. Figure Eight. Available online: <https://www.figure-eight.com/platform/> (accessed on 7 March 2020).
9. Lionbridge. Available online: <https://lionbridge.ai/services/crowdsourcing/> (accessed on 7 March 2020).
10. De Winter, J.; Kyriakidis, M.; Dodou, D.; Happee, R. Using CrowdFlower to study the relationship between self-reported violations and traffic accidents. *Procedia Manuf.* **2015**, *3*, 2518–2525. [CrossRef]
11. Vaughan, J.W. Making better use of the crowd: How crowdsourcing can advance machine learning research. *J. Mach. Learn. Res.* **2017**, *18*, 7026–7071.
12. Li, X.; Shi, W.; Zhu, B. The face of internet recruitment: Evaluating the labor markets of online crowdsourcing platforms in China. *Res. Politics* **2017**, *5*, 1–8. [CrossRef]
13. Alotaibi, B.; Abbasi, R.A.; Aslam, M.A.; Saeedi, K.; Alahmadi, D. Startup Initiative Response Analysis (SIRA) Framework for Analyzing Startup Initiatives on Twitter. *IEEE Access* **2020**, *8*, 10718–10730. [CrossRef]
14. Alshutayri, A.; Atwell, E. Creating an Arabic dialect text corpus by exploring Twitter, Facebook, and online newspapers. In *OSACT 3 Proceedings. OSACT 3 the 3rd Workshop on Open-Source Arabic Corpora and Processing Tools, Co-Located with LREC 2018, Miyazaki, Japan, 8 May 2018*; LREC: Pelican Rapids, MN, USA, 2018.
15. Alsarsour, I.; Mohamed, E.; Suwaileh, R.; Elsayed, T. Dart: A large dataset of dialectal Arabic tweets. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan, 7–12 May 2018.
16. Al-Twairish, N.; Al-Matham, R.; Madi, N.; Almugren, N.; Al-Aljmi, A.-H.; Alshalan, S.; Alshalan, R.; Alrumayyan, N.; Al-Manea, S.; Bawazeer, S. Suar: Towards building a corpus for the Saudi dialect. *Procedia Comput. Sci.* **2018**, *142*, 72–82. [CrossRef]
17. Mubarak, H. Crowdsourcing Speech and Language Data for Resource-Poor Languages. In *Proceedings of the International Conference on Advanced Intelligent Systems and Informatics, Cairo, Egypt, 9–11 September 2017*; Springer: Berlin/Heidelberg, Germany, 2017.
18. Alhelbawy, A.; Massimo, P.; Kruschwitz, U. Towards a corpus of violence acts in arabic social media. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, Portorož, Slovenia, 23–28 May 2016.
19. Zaidan, O.F.; Callison-Burch, C. Arabic dialect identification. *Comput. Linguist.* **2014**, *40*, 171–202. [CrossRef]

20. Albadi, N.; Kurdi, M.; Mishra, S. Investigating the effect of combining GRU neural networks with handcrafted features for religious hatred detection on Arabic Twitter space. *Soc. Netw. Anal. Min.* **2019**, *9*, 41. [CrossRef]
21. Bougrine, S.; Cherroun, H.; Abdelali, A. Altruistic crowdsourcing for arabic speech corpus annotation. *Procedia Comput. Sci.* **2017**, *117*, 137–144. [CrossRef]
22. Akila, G.; El-Menisy, M.; Khaled, O.; Sharaf, N.; Tarhony, N.; Abdennadher, S. Kalema: Digitizing arabic content for accessibility purposes using crowdsourcing. In Proceedings of the International Conference on Intelligent Text Processing and Computational Linguistics, Cairo, Egypt, 14–20 April 2015; Springer: Berlin/Heidelberg, Germany, 2015.
23. Almeman, K.; Lee, M. Automatic building of arabic multi dialect text corpora by bootstrapping dialect words. In Proceedings of the 2013 1st International Conference on Communications, Signal Processing, and Their Applications (ICCSIPA), Sharjah, UAE, 12–14 February 2013; IEEE: New York, NY, USA, 2013.
24. Baly, R.; Khaddaj, A.; Hajj, H.; El-Hajj, W.; Shaban, K.B. Arsentd-lev: A multi-topic corpus for target-based sentiment analysis in arabic levantine tweets. *arXiv* **2019**, arXiv:1906.01830.
25. Farra, N.; McKeown, K.; Habash, N. Annotating targets of opinions in arabic using crowdsourcing. In Proceedings of the Second Workshop on Arabic Natural Language Processing, Beijing, China, 30 July 2015.
26. Al-Muzaini, H.A.; Al-Yahya, T.N.; Benhidour, H. Automatic Arabic Image Captioning using RNN-LSTM-Based Language Model and CNN. *Int. J. Adv. Comput. Sci. Appl.* **2018**, *9*. [CrossRef]
27. Tashkeel Crowdsourcing Platform [تشكيل]. Available online: <http://crowdsourcing.kauwld.com/> (accessed on 3 September 2020).
28. Öhlen, J.; Morse, J.M.; Niehaus, L. Mixed Method Design: Principles and Procedures. In *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research*; Routledge: London, UK, 2010.
29. Denzin, N.K.; Lincoln, Y.S. Introduction: The discipline and practice of qualitative research. In *The Sage Handbook of Qualitative Research*; Sage Publications Ltd.: New York, NY, USA, 2008.
30. Lieberman, E.S. Nested analysis as a mixed-method strategy for comparative research. *Am. Political Sci. Rev.* **2005**, *99*, 435–452. [CrossRef]
31. Brooke, J. *A Quick and Dirty Usability Scale.* *Usability Evaluation in Industry*; Taylor & Francis Ltd.: London, UK, 1996.
32. Brooke, J. SUS: A retrospective. *J. Usability Stud.* **2013**, *8*, 29–40.
33. Sauro, J. Measuring Usability with the System Usability Scale (SUS). 2011. Measuring U, Blog, 2 February 2011. Available online: www.measuringu.com (accessed on 6 September 2020).
34. Lewis, J.R.; Sauro, J. The factor structure of the system usability scale. In Proceedings of the International Conference on Human Centered Design, San Diego, CA, USA, 19–24 July 2009; Springer: Berlin/Heidelberg, Germany, 2009.
35. Amazon Mechanical Turk. Available online: <https://www.mturk.com/> (accessed on 15 July 2020).
36. Appen. Available online: <https://appen.com/> (accessed on 15 July 2020).
37. Techopedia—IT Dictionary for Computer Terms and Tech Definitions. Available online: <https://www.techopedia.com/dictionary> (accessed on 16 December 2019).



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).