KSAA-CAD Shared Task: Contemporary Arabic Dictionary for Reverse Dictionary and Word Sense Disambiguation

Waad Alshammari, Amal Almazrua, Asma Al Wazrah, Rawan Al-Matham, Muneera Alhoshan, Abdulrahman AlOsaimy, Afrah Altamimi, Abdullah Alfaifi
King Salman Global Academy for Arabic Language
{walshammari,aalmazrua,aalwazrah,ralmatham, malhoshan,aalosaimy,a.altamimi,aalfaifi}@ksaa.gov.sa

Abstract

This paper outlines the KSAA-CAD shared task, highlighting the Contemporary Arabic Language Dictionary within the scenario of developing a Reverse Dictionary (RD) system and enhancing Word Sense Disambiguation (WSD) capabilities. The first KSAA-RD (Al-Matham et al., 2023) highlighted significant gaps in the domain of RDs, which are designed to retrieve words by their meanings or definitions. This shared task comprises two tasks: RD and WSD. The RD task focuses on identifying word embeddings that most accurately match a given definition, termed a "gloss". Conversely, the WSD task involves determining the specific meaning of a word in context, particularly when the word has multiple meanings. The winning team achieved the highest-ranking score of 0.0644 in RD using Electra embeddings. However, the baseline still surpasses all participant models in terms of rank, while the Asos team achieved the best score in terms of cosine similarity and mean squared error. In this paper, we describe the methods employed by the participating teams and provide insights into the future direction of KSAA-CAD.

1 Introduction

The KSAA-CAD is considered an extension of the first KSAA-RD iteration (Al-Matham et al., 2023). These two shared tasks utilized Contemporary Arabic Language Dictionaries (CAD). While the first iteration was primarily focused on the Reverse Dictionary (RD) with two subtasks, Arabic RD and Cross-Lingual RD (CLRD), the current iteration extends its scope to include Word Sense Disambiguation (WSD) task. The second RD task includes multiple definition from different dictionaries, while the first one includes definitions from only one dictionary.

RD enables users to find words by describing their meanings or definitions, rather than searching by exact word as in traditional dictionaries. Users input a description or phrase, and the RD provides a list of corresponding words. On the other hand, the objective of the WSD task is to assist systems in determining the most appropriate sense of new terms and definitions, emphasizing the importance of context in understanding the intended sense.

This shared task is conducted as a part of the second Arabic Natural Language Processing (ArabicNLP) conference, held in conjunction with ACL 2024. It featured the KSAA-CAD (King Salman Global Academy for Arabic Language).

Four research papers were submitted for the KSAA-CAD shared task. The datasets and codes developed for the KSAA-CAD shared task are publicly accessible in a GitHub repository [1], contributing to ongoing Arabic NLP research efforts.

This paper is structured into seven sections. Section 2 delves into related work, section 3 presents the data for the shared task, and section 4 defines the KSAA-CAD tasks. Section 5 elucidates the performance evaluation metric. Finally, section 6 discusses the baseline model and the results achieved by the participating teams.

2 Related work

2.1 Reverse dictionary

In our previous work (Al-Matham et al., 2023) presented the KSAA-RD shared task to develop a RD system for Arabic. The task includes two subtasks: Arabic RD and CLRD. The approach involves generating word embeddings from definitions (glosses) in Arabic and English.

[1] https://github.com/ksaa-nlp/KSAA-CAD

Proceedings of The Second Arabic Natural Language Processing Conference, pages 677–685
August 16, 2024 ©2024 Association for Computational Linguistics

Participating teams employed various methods, focusing on leveraging neural language models and embedding techniques to achieve high accuracy in identifying corresponding words from provided definitions.

Since the last edition, the field of RD systems has made significant progress. Notably, the work by (Tian et al., 2024a) explores the use of large-scale language models (LLMs) through prompt engineering to solve the RD problem. Their approach involves a small-scale language model that generates candidate words by analyzing definitions across multiple semantic dimensions and incorporating negative samples. These candidate sets are then used to construct effective prompts, enhancing the performance of larger LLMs without the need for extensive computational resources. This method highlights the effective combination of small and large models to optimize RD functionalities.

Another work by (Tian et al., 2024b) introduces the RDMTL model, which employs multitask learning to address the challenge of distinguishing between words with similar definitions. The RDMTL model integrates a primary task for extracting semantic features from definitions with auxiliary tasks for part-of-speech tagging and sentence generation. By utilizing a Bi-LSTM structure enhanced with CNN and attention mechanisms, RDMTL effectively captures multilevel semantic information, addressing subtle differences between words with similar meanings and improving overall model performance.

The literature presents various methods for developing RDs, including the Neural Language Model-based strategies mentioned by (Agrawal et al., 2021; Hedderich et al., 2019; Hill et al., 2016; Morinaga and Yamaguchi, 2018; Morinaga and Yamaguchi, 2020; Pilehvar, 2019; Qi et al., 2020; Yan et al., 2020; Zhang et al., 2020). However, until now, the Arabic RD problem hasn't been solved yet. The only efforts have been the proposed are previous efforts in KSAA- shared task.

In this year's shared task, we provide a new dataset with a wider range of vocabulary and complex sentence structures that aim to push the boundaries of current Arabic RD systems.

2.2 Word Sense Disambiguation

WSD is the process of accurately identifying the specific sense or meaning of a word, which is referred to as "gloss" in this study, by examining its surrounding context. This area has received considerable development and attention in the English language domain. This demonstrates the significant challenge in accurately determining the meaning of Arabic words based on their usage in varying contexts. Nonetheless, the progress in the application of WSD techniques to Arabic lags behind. The challenge of word ambiguity in Arabic is especially significant. About 43% of Arabic words with diacritics carry multiple meanings (Debili et al., 2002; El-Razzaz et al., 2021), and this figure rises sharply to 72% in the case of words without diacritics, underscoring the intricacy of interpreting Arabic script (Alqahtani et al., 2019). The development of Arabic WSD is hindered by the scarcity of datasets annotated for sense. Moreover, the inherent complexity of WSD is compounded by the semantic polysemy prevalent in Arabic words, as noted by (Al-Hajj and Jarrar, 2021). The Arabic word "علم" exemplifies the complexity of WSD. In the context of "رفرف يا علم", the word "عَلَم" refers to a 'flag' (راية، ولواء). In a different context, such as "زيد علم على شخص عاقل مذكر", the word "عَلَم" denotes a 'name' (هو الاسم الذي يدل على الذي سُمي به ...). Meanwhile, in the context "لا ينال العلم إلا فتى ..." the word "عِلْم" signifies 'knowledge' (معرفة، ودراية).

3 Dataset

This section examines the data employed in KSAA-CAD for both RD and WSD tasks.

3.1 Reverse dictionary

The RD data is organized into two main components: the dictionary data and the word embedding vectors. For generating these word embeddings, our approach is to utilize three distinct architectures of contextualized word embedding.

3.1.1 Dictionary data

In the first iteration of KSAA-RD (Al-Matham et al., 2023), the dataset derived from a single source: the "Contemporary Arabic Language Dictionary" by Ahmed Mokhtar Omar (Omar, 2008). In this revised edition, we endeavor to expand our sources to encompass three dictionaries of Contemporary Arabic Language. The first of these is the newly released "Alriyadh Dictionary" (KSAA, 2023). The second is the "Contemporary Arabic Language Dictionary" by Ahmed Mokhtar Omar (Omar, 2008), a resource previously utilized in the first iteration KSAA-RD. The third is the "Al Wassit LMF Arabic Dictionary" (Namly, 2015).

The three dictionaries employ the transferred version of this lexicon which conforms to the ISO standard, specifically the Lexical Markup Framework (LMF) (Aljasim et al., 2022). The dictionary is based on lemmas rather than roots. These dictionaries comprise words, commonly referred to as lemmas, and these may come with glosses, part of speech (POS).

We utilized a subset of the Alriyadh Dictionary, which contains 10K entries with 8,909 unique lemmas. Our specific focus is Our specific focus was on identifying unique lemmas that have corresponding entries in both the Ahmed Mokhtar Omar and Al Wassit dictionaries. Table 2 provides details of these selected entries.

Dictionary	Entry
Alriyadh Dictionary	10,626
Ahmed Mokhtar Omar	3,341
Al Wassit LMF Arabic Dictionary	25,248
All dictionaries	39,215

Table 1: Dictionaries Statistic.

3.1.2 Embedding data

Experiments conducted on the first iteration of KSAA-RD (Al-Matham et al., 2023) revealed that fixed word embedding representations such as word2vec (Mikolov et al., 2013; Soliman et al., 2017) did not yield satisfactory performance. Consequently, in this edition, our focus shifted to contextualized word embeddings, which have demonstrated improved performance in KSAA-RD. Accordingly, we utilized advanced models such as Electra (Clark et al., 2020) and BERT (Devlin et al., 2019), to enhance the effectiveness of the system. Specifically, our objective is to employ AraELECTRA (Antoun et al., 2021), AraBERTv2 (Antoun et al., 2020), and camelBERT-MSA (Inoue et al., 2021), referred to respectively as Electra, bertSEG, and bertMSA. AraELECTRA, an Arabic language representation model, is developed based on the Electra framework. Instead of training the model to recover masked tokens, Electra is designed to train a discriminator model. AraBERTv2 and camelBERT-MSA are both Arabic language models developed based on BERT architecture. The latter, camelBERT-MSA, is pretrained on a Modern Standard Arabic (MSA) corpus.

3.2 Word Sense Disambiguation

The dataset itself comprises two core components: the WSD context gloss mapping data and dictionary data. The WSD context gloss mapping consists of word, context, context ID, and corresponding gloss ID. As a concrete instance, Figure 1 (d) depicts an example from the training dataset for the corresponding WSD JSON.

The dictionary data contains word, gloss, and gloss ID. The dictionary data is derived from the "Contemporary Arabic Language Dictionary" by Ahmed Mokhtar Omar (Omar, 2008). As a concrete instance, Figure 1 (e) depicts an example from the WSD dictionary.

3.3 Dataset description

The RD and WSD datasets are both in JSON format. The RD dataset contains approximately 39K entries in Contemporary Arabic Language, while the WSD dataset features around 28K entries. Both datasets are thoughtfully divided into three sections: a training split comprising 80%, a validation split representing 10%, and a test split representing 10% of the data points. Refer to Table 1 for data statistics.

Task	Train	Dev	Test
RD	31,372	3,921	3,922
WSD	22,404	2,801	2,801
WSD dictionary	15,865		

Table 2: Dataset Statistics.

4 Task Description

This section provides a detailed overview of the two tasks: RD and WSD. The RD task aims to transform the Arabic word definition or gloss into embeddings. Meanwhile, the WSD task focuses on determining the intended meaning or sense of a word in the context of a given sentence or segment.

4.1 Task 1: Reverse Dictionary

RDs, identified by their sequence-to-vector format, introduce a differentiated strategy in contrast to traditional dictionary lookup methods. The RD task concentrates on the conversion of human-readable glosses into word embedding vectors.

This process entails reconstructing the word embedding vector corresponding to the defined word, a methodology aligning with the approaches of (Mickus et al., 2022; Zanzotto et al., 2010; Hill et al., 2016).

The dataset includes lemma, lemma vector representations, and their respective gloss, as depicted in Figure 1, parts (a) and (b). The developed model is expected to generate novel lemma vector representations for the unseen human-readable definitions in the test set. This strategy enables users to search for words based on anticipated definitions or meanings.

4.2 Task 2: Word Sense Disambiguation

WSD focuses on identifying the specific sense of a word in a given context. The WSD gloss-based approach is categorized as a knowledge-based WSD method. This approach utilizes external resources, especially dictionaries. This technique involves determining a word's intended meaning by calculating the overlap between its contextual use and the provided gloss or definition.

In the realm of contemporary Arabic language, dictionaries have been utilized in the development of gloss-based WSD datasets, as evidenced in the works of (Jarrar et al., 2023; El-Razzaz et al., 2021). These studies employed the Ahmed Mokhtar Omar dictionary (Omar, 2008). Furthermore, the research conducted by (Jarrar et al., 2023) also incorporated the Al-Ghani Al-Zaher dictionary (Abul-Azm, 2014).

Following (Jarrar et al., 2023; El-Razzaz et al., 2021), the gloss-based WSD is structured as a sentence-pair binary classification problem.

Our objective is to provide a dataset comprising 40K context-gloss pairs.

This dataset includes context, sense, lemma, sense vector representations, and sense-specific IDs, as illustrated in Figure 1, parts (c) and (d). The developed model is designed to calculate the semantic overlap between unseen context and human-readable gloss in the test set. This approach enables users to disambiguate words based on their anticipated definitions or meanings.

5 Evaluation

The main objective of the RD task lies in the reconstruction of word embeddings. Subsequent to the first KSAA-RD task, our methodology incorporated three approaches for evaluating vector similarity. These include the Mean Squared Error (MSE), Cosine Similarity measurement, and

id	**ar.78**
word	**عين**
POS	**n**
gloss	**نبع الماء**

(a) Example of RD task data point

```
{
  "id":"ar.78",
  "word":"عين",
  "gloss":"...نبع الماء",
  "pos":"n",
  "electra":[0.7, 0.3, …],
  "arabert":[0.2, 0.8, …],
  "camelbert":[1.3, 0.5, …],

 }
```

(b) Corresponding RD JSON snippet

context_id	**context.301**
Context	**يأتي برمجان اللغة العربية...**
word	**اللغة**
gloss	**كُلُّ وسيلة لتبادل المشاعر والأفكار كالإشارات**

(c) Example of WSD task data point

```
{

"context_id":"context.301",
"context":" يأتي برمجان اللغة
العربية...",

"word": "اللغة",
"gloss_id":"gloss.205",
"lemma_id":"ar.301",

}
```

(d) Corresponding WSD JSON snippet

```
{
"lemma_id": "ar.301",
"gloss_id":"gloss.205",
"gloss":" كُلُّ وسيلة لتبادل المشاعر
والأفكار كالإشارات ..."
}
{
"lemma_id": "ar.301",
"gloss_id":"gloss.211",
"gloss":" علم يختص بدراسة اللُّغة
دراسة منهجيَّة في إطار..."
}
```

(e) WSD dictionary

Figure 1: The structure of a data point.

Ranking Evaluation Metrics, as proposed in the CODWOE SemEval competition (Mickus et al., 2022). The ranking metric can be described as follows:

$$\text{Ranking}(p_i) = \frac{\sum_{t_j\, Test\, set} 1_{\cos(p_i,t_j) > \cos(p_i,t_i)}}{\#\, Test\ set} \quad (1)$$

On the other hand, Accuracy and MRR@2 (Mean Reciprocal Rank at 2) are the primary metrics employed in the WSD task. These metrics are crucial as they provide a measure of how accurately the system identifies the correct sense of a word within its context. Accuracy represents the proportion of correct predictions made by the model, while MRR@2 measures the average rank position of the correct answer, rewarding models that place the correct answer higher in the ranked list.

6 Shared Task Teams & Results

This section introduces baseline models, outline the participating teams, and provide an overview of the submitted systems along with their results.

6.1 Our Baseline system

For the RD task, we utilize SOTA MARBERT (Abdul-Mageed et al., 2021) and CamelBERT-MSA (Inoue et al., 2021) models, employing fine-tuning techniques to excel in Arabic RD. These models are evidenced by their superior performance in the KSAA-RD shared task (Al-Matham et al., 2023), marking them as winning approaches. Notably, our data preprocessing involves removing punctuation, special characters, diacritics, and non-Arabic letters.

For the WSD task, dataset is enriched using lemma IDs by joining WSD entries with the WSD dictionary to incorporate both relevant and irrelevant glosses. After cleaning the data, the model is trained to determine the relevance of a gloss to a word in context. The highest probability gloss is then calculated for each word in context, improving its ability to accurately identify context-appropriate meanings. We employ two approaches for WSD:

- **Fine-tuning:** The approach leverages *BertForSequenceClassification*, specifically with CamelBERT-MSA and AraBERTv2. The target word in context is wrapped with special tokens "<token>word</token>". CamelBERT-MSA and AraBERTv2 achieved accuracy rates of 91.61% and 91.25%, respectively.
- **Neural Network:** This approach involves feeding the three text embeddings (context, word, and gloss) from the multilingual-E5-base model into a simple LSTM neural network consisting of an input layer, a single LSTM layer, a dense layer, and an output layer. With this E5+LSTM model, an accuracy of 88.83% is achieved.

6.2 Participating Teams

A total of 46 teams registered, resulting in 90 valid submissions from 5 different teams. During testing, 5 submissions were received for the RD task, and only 1 for WSD. However, WSD was excluded due to the team's performance being 64% lower than the baseline. Furthermore, 4 description papers were submitted and accepted. Detailed information about these 4 teams can be found in Table 3.

Team	Affiliation
ASOS	Prince Sultan University RIOTU Lab
MISSION	N/A
Baleegh	Princess Nourah bint Abdulrahman University
Cher	University of Edinburgh and University of Cambridge

Table 3: List of participating teams

6.3 Results and Description of Submitted Systems

Four teams participated in the RD task: ASOS team, Baleegh team, MISSION team, and Cher team. Notably, all teams utilized transformer-based models. The results for the RD tasks are presented in Table 4.

ASOS (Sibaee et al., 2024) propose the use of a semi-encoder with four hidden layers for the Arabic RD task. They begin by expanding the dataset using last year's task data, utilizing only ArElectra embeddings. Then they explore multiple sentence transformer models alongside the AraBERTv2 model. The generated embeddings are fed into the semi-encoder to produce the final result. Their system achieved the best rank and MSE score of 0.0644 and 0.059 using the Electra model with AraBERTv2 embeddings on the test set compared with other teams, for rank and MSE score respectively. However, the baseline score is still better than their score in terms of rank. Although they achieved the best result, they trained their final model on both (the training and development datasets) which enhances their results.

Cher (Chen et al., 2024) employ a multi-task learning framework to enhance the Arabic RD task. The authors developed a model that jointly learns RD, definition generation, and reconstruction tasks using transformer layers with GeLU activation. They explored different tokenization strategies, including whitespace, Farasa segmentation, and CAMeLBERT, to process definitions. The model embeds words and definitions into a shared space, optimized through a bottleneck design. The training was conducted in a data-constrained environment using provided embeddings without external resources. Evaluation results demonstrated the model's effectiveness, with the AraBERT (Farasa) tokenizer yielding a promising performance across metrics in the development phase, with a result of 0.4834 in the ranking score for bertMSA in the test phase.

Baleegh (Alheraki and Meshoul, 2024) employ the AraT5 V2 (Nagoudi et al., 2021) model for the Arabic RD task. Their methodology begins with preparing a new dataset using a data enrichment technique that combines glosses with contextually relevant examples from Arabic Wikipedia embeddings. The datasets are then tokenized using SentencePiece tokenizer. Subsequently, the input is fed into the AraT5 V2 encoder, followed by a pooling layer and a linear layer to produce the final prediction. Their methodology achieved the highest RANK score of 0.1781 with Electra embedding with gloss only on test set. It stated that adding examples to the gloss did not improve the outcomes.

MISSION (Alharbi, 2024) employs the same model as (Alheraki and Meshoul, 2024) which utilize AraT5 V2 model for the Arabic RD task. The approach involves fine-tuning the model for predicting word embeddings based on input glosses. The model architecture includes a T5 structure followed by a linear layer for embedding generation. Training involves providing tokenized input sequences, attention masks, and target word tokens. The methodology utilized gloss-to-embedding mapping, with a ranking score of 0.2482 for Electra in the test set.

The findings for the RD task reveal the performance of four participating teams. The evaluation utilized three types of contextualized embeddings: Electra, bertSEG, and bertMSA. The ASOS team achieved the highest-ranking score with 0.0644 using Electra embeddings, though their scores for bertSEG 0.198 and bertMSA 0.1484 were lower. It was observed that two papers employed the AraT5 V2 model, while one paper utilized the AraBERTv2 model. Additionally, one model incorporated transformer layers in its architecture.

From the previous KSAA-RD shared task, we noticed that the Electra embedding had the best performance for the winning selections. Additionally, in this iteration, Electra continues to showcase exceptional performance across all team models compared to other embeddings. Therefore, it can be utilized to enhance the results for this problem. Also, it's noteworthy that both CamelBERT and MARBERT baselines, leveraging all embeddings, particularly regarding the rank metric, surpassed all participant models, marking a significant advancement in state-of-the-art results. This underscores the varied behavior of each model with respect to different contextualized embedding representations.

On the other hand, in terms of the cosine similarity metric, the ASOS team achieved the best result using the bertSEG embedding. Additionally, the MISSION and Baleegh teams achieved a comparable result, with almost a 6% difference, also utilizing bertSEG. This underscores the significant impact of bertSEG on cosine similarity metrics, a finding that is also highlighted in our baselines.

7 Conclusion

In this paper, we introduced the KSAA-CAD shared task, a thorough effort focused on progressing RD systems and WSD in Arabic. This work expands the initial version of KSAA-RD by incorporating new datasets and integrating modern embedding models such as Electra, AraBERTv2, and camelBERT-MSA.

Notably, for all participant teams, Electra achieved superior results compared to other embeddings using the same model. On the other hand, bertSEG surpassed others in terms of cosine similarity performance. Nevertheless, there is still potential for enhancement, particularly in utilizing other approaches such as LLM and prompt engineering (Tian et al., 2024a).

In conclusion, the RD baseline surpasses all participant models in terms of rank. This achievement showcases how effectively contextualized word embeddings in handling Arabic RD tasks. Additionally, it clarifies that the Arabic RD problem requires further investigation. Asos team has conducted an analysis of the datasets that can be used in the future to aid in RD model training.

In the WSD task, gloss-based methods showed potential in the baseline result but were not as effective as RD because of the complexity of Arabic word meanings and the lack of annotated datasets. It is crucial to emphasize the team's low performance in the WSD task, which was 64% below the baseline. This highlights significant obstacles in the current methods of WSD in Arabic, emphasizing the need for more comprehensive datasets, as well as more advanced models, to effectively address the complexities of Arabic word meanings.

The involvement and outcomes of different teams offered valuable perspectives on the techniques that can be used for future progress. Groups that employed transformer-based models and creative data augmentation methods, like incorporating Wikipedia embeddings and utilizing semi-encoder architectures, showcased significant enhancements in performance.

	Embedding	Dev			Test		
		Cos↑	MSE↓	Rank↓	Cos↑	MSE↓	Rank↓
Baseline CamelBERT	Electra	0.7368	0.1458	0.0084	0.5065	0.2459	0.0334
	bertSEG	0.8436	0.0555	0.0126	**0.7556**	0.0831	0.0334
	bertMSA	0.8185	0.2195	0.0110	0.6965	0.3477	**0.0330**
Baseline MARBERT	Electra	0.5132	0.2436	0.0335	0.5104	0.2444	0.0334
	bertSEG	0.7604	0.0818	0.0335	0.7610	0.0816	0.0334
	bertMSA	0.6949	0.3495	0.0335	0.6970	0.3473	0.0334
ASOS	Electra	---	---	---	0.7071	0.158	**0.0644**
	bertSEG	---	---	---	**0.8300**	**0.059**	0.198
	bertMSA	---	---	---	0.807	0.229	0.1484
MISSION	Electra	0.5515	0.2297	0.2469	0.5507	0.2298	0.2482
	bertSEG	0.7745	0.0777	0.4126	**0.7731**	0.0781	0.4165
	bertMSA	0.7188	0.3257	0.3334	0.7219	0.3224	0.3315
Baleegh	Electra	0.5686	0.2255	0.1721	0.5678	0.2257	0.1781
	bertSEG	0.7752	0.0776	0.3518	**0.7739**	0.0779	0.3522
	bertMSA	0.7140	0.3330	0.3039	0.7168	0.3299	0.3021
Cher	Electra	---	---	---	---	---	---
	bertSEG	---	---	---	0.7012	0.3446	0.4913
	bertMSA	---	---	---	0.7671	0.0800	0.4834

Table 4: Participants' Results for RD task

References

Muhammad Abdul-Mageed, AbdelRahim A. Elmadany, and El Moatez Billah Nagoudi. 2021. ARBERT & MARBERT: Deep Bidirectional Transformers for Arabic. CoRR, abs/2101.01785. arXiv: 2101.01785.

Abdul-Ghani Abul-Azm. 2014. Al-ghani al-zaher dictionary.

Aarchi Agrawal, Athulkumar R, K S Ashin Shanly, Kavita Vaishnaw, and Mayank Singh. 2021. Reverse Dictionary Using an Improved CBoW Model. In Proceedings of the 3rd ACM India Joint International Conference on Data Science & Management of Data (8th ACM IKDD CODS & 26th COMAD), page 420, New York, NY, USA. Association for Computing Machinery.

Moustafa Al-Hajj and Mustafa Jarrar. 2021. ArabGlossBERT: Fine-Tuning BERT on Context-Gloss Pairs for WSD. In Ruslan Mitkov and Galia Angelova, editors, Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2021), pages 35–43, Held Online. INCOMA Ltd.

THAMER MASEER Alharbi. 2024. MISSION at KSAA-CAD 2024: AraT5 with Arabic Reverse Dictionary. In Proceedings of the 2nd Arabic Natural Language Processing Conference (Arabic-NLP), Part of the ACL 2024.

Mais Alheraki and Souham Meshoul. 2024. Baleegh at KSAA-CAD 2024: Towards Enhancing Arabic Reverse Dictionaries. In Proceedings of the 2nd Arabic Natural Language Processing Conference (Arabic-NLP), Part of the ACL 2024. Association for Computational Linguistics.

Hawra Aljasim, Sultanah Alghurabi, Halah Alharbi, Abdulrahman Alosaimy, Afrah Altamimi, Asyah Almutawa, and Abdullah Alfifi. 2022. Toward an Automatic Semantic Order of Word Meanings Based on Lexically Tagged Corpus. Manuscript submitted for publication.

Rawan Al-Matham, Waad Alshammari, Abdulrahman AlOsaimy, Sarah Alhumoud, Asma Wazrah, Afrah Altamimi, Halah Alharbi, and Abdullah Alaifi. 2023. KSAA-RD Shared Task: Arabic Reverse Dictionary. In Hassan Sawaf, Samhaa El-Beltagy, Wajdi Zaghouani, Walid Magdy, Ahmed Abdelali, Nadi Tomeh, Ibrahim Abu Farha, Nizar Habash, Salam Khalifa, Amr Keleg, Hatem Haddad, Imed Zitouni, Khalil Mrini, and Rawan Almatham, editors, Proceedings of ArabicNLP 2023, pages 450–460, Singapore (Hybrid).

Sawsan Alqahtani, Hanan Aldarmaki, and Mona Diab. 2019. Homograph Disambiguation through Selective Diacritic Restoration. In Wassim El-Hajj, Lamia Hadrich Belguith, Fethi Bougares, Walid Magdy, Imed Zitouni, Nadi Tomeh, Mahmoud El-Haj, and Wajdi Zaghouani, editors, Proceedings of the Fourth Arabic Natural Language Processing Workshop, pages 49–59, Florence, Italy.

Wissam Antoun, Fady Baly, and Hazem Hajj. 2020. AraBERT: Transformer-based Model for Arabic Language Understanding. In Proceedings of the 4th Workshop on Open-Source Arabic Corpora and Processing Tools, with a Shared Task on Offensive Language Detection, pages 9–15, Marseille, France. European Language Resource Association.

Wissam Antoun, Fady Baly, and Hazem Hajj. 2021. AraELECTRA: Pre-Training Text Discriminators for Arabic Language Understanding. In Proceedings of the Sixth Arabic Natural Language Processing Workshop, pages 191–195, Kyiv, Ukraine (Virtual). Association for Computational Linguistics.

Pinzhen Chen, Zheng Zhao, and Shun Shao. 2024. Cher at KSAA-CAD 2024: Compressing Words and Definitions into the Same Space for Arabic Reverse Dictionary. In Proceedings of the 2nd Arabic Natural Language Processing Conference (Arabic-NLP), Part of the ACL 2024. Association for Computational Linguistics.

Kevin Clark, Minh-Thang Luong, Quoc V. Le, and Christopher D. Manning. 2020. ELECTRA: Pre-training Text Encoders as Discriminators Rather Than Generators. CoRR, abs/2003.10555. arXiv: 2003.10555.

Fathi Debili, Hadhemi Achour, and Emna Souissi. 2002. La langue arabe et l'ordinateur de l'étiquetage gramatical à la voyellation automatique. In

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of the 2019 Conference of the North, pages 4171–4186, Minneapolis, Minnesota.

Mohammed El-Razzaz, Mohamed Waleed Fakhr, and Fahima A. Maghraby. 2021. Arabic Gloss WSD Using BERT. Applied Sciences, 11(6).

Michael A. Hedderich, Andrew Yates, Dietrich Klakow, and Gerard de Melo. 2019. Using Multi-Sense Vector Embeddings for Reverse Dictionaries. In Proceedings of the 13th International Conference on Computational Semantics - Long Papers, pages 247–258, Gothenburg, Sweden.

Felix Hill, Kyunghyun Cho, Anna Korhonen, and Yoshua Bengio. 2016. Learning to Understand Phrases by Embedding the Dictionary. Transactions of the Association for Computational Linguistics, 4:17–30.

Go Inoue, Bashar Alhafni, Nurpeiis Baimukan, Houda Bouamor, and Nizar Habash. 2021. The Interplay of Variant, Size, and Task Type in Arabic Pre-trained Language Models. In Proceedings of the Sixth Arabic Natural Language Processing Workshop, pages 92–104, Kyiv, Ukraine (Virtual).

Mustafa Jarrar, Sanad Malaysha, Tymaa Hammouda, and Mohammed Khalilia. 2023. SALMA: Arabic Sense-Annotated Corpus and WSD Benchmarks. In Hassan Sawaf, Samhaa El-Beltagy, Wajdi Zaghouani, Walid Magdy, Ahmed Abdelali, Nadi Tomeh, Ibrahim Abu Farha, Nizar Habash, Salam Khalifa, Amr Keleg, Hatem Haddad, Imed Zitouni, Khalil Mrini, and Rawan Almatham, editors, Proceedings of ArabicNLP 2023, pages 359–369, Singapore (Hybrid).

king Salman global academy for Arabic language KSAA. 2023. Alriyadh Dictionary.

Timothee Mickus, Kees Van Deemter, Mathieu Constant, and Denis Paperno. 2022. Semeval-2022 Task 1: CODWOE – Comparing Dictionaries and Word Embeddings. In Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval-2022), pages 1–14, Seattle, United States. Association for Computational Linguistics.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient Estimation of Word Representations in Vector Space. arXiv:1301.3781 [cs].

Yuya Morinaga and Kazunori Yamaguchi. 2018. Improvement of Reverse Dictionary by Tuning Word Vectors and Category Inference. In Robertas Damaševičius and Giedrė Vasiljevienė, editors, Information and Software Technologies, pages 533–545, Cham. Springer International Publishing.

Yuya Morinaga and Kazunori Yamaguchi. 2020. Improvement of Neural Reverse Dictionary by Using Cascade Forward Neural Network. Journal of Information Processing, 28:715–723.

El Moatez Billah Nagoudi, AbdelRahim A. Elmadany, and Muhammad Abdul-Mageed. 2021. AraT5: Text-to-Text Transformers for Arabic Language Understanding and Generation. CoRR, abs/2109.12068. arXiv: 2109.12068.

Driss Namly. 2015. “Al wassit” LMF Arabic dictionary.

Ahmed Mukhtar Omar. 2008. Dictionary of Contemporary Arabic Language. Volume One, 1st Edition, Cairo, Egypt:2109.

Mohammad Taher Pilehvar. 2019. On the Importance of Distinguishing Word Meaning Representations: A Case Study on Reverse Dictionary Mapping. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), pages 2151–2156, Minneapolis, Minnesota.

Fanchao Qi, Lei Zhang, Yanhui Yang, Zhiyuan Liu, and Maosong Sun. 2020. WantWords: An Open-source Online Reverse Dictionary System. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, pages 175–181, Online.

Serry Sibaee, Abdullah I. Alharbi, Samar Ahmad, OMER NACAR, and Anis Koubaa. 2024. ASOS at KSAA-CAD 2024: One Embedding is All You Need for Your Dictionary. In Proceedings of the 2nd Arabic Natural Language Processing Conference (Arabic-NLP), Part of the ACL 2024.

Abu Bakr Soliman, Kareem Eissa, and Samhaa R. El-Beltagy. 2017. AraVec: A set of Arabic Word Embedding Models for use in Arabic NLP. Procedia Computer Science, 117:256–265.

Sicheng Tian, Shaobin Huang, Rongsheng Li, and Chi Wei. 2024a. A prompt construction method for the reverse dictionary task of large-scale language models. Engineering Applications of Artificial Intelligence, 133:108596.

Sicheng Tian, Shaobin Huang, Rongsheng Li, Chi Wei, and Ye Liu. 2024b. RDMTL: Reverse dictionary model based on multitask learning. Knowledge-Based Systems, 296:111869.

Hang Yan, Xiaonan Li, Xipeng Qiu, and Bocao Deng. 2020. BERT for Monolingual and Cross-Lingual Reverse Dictionary. In Findings of the Association for Computational Linguistics: EMNLP 2020, pages 4329–4338, Online.

Fabio Massimo Zanzotto, Ioannis Korkontzelos, Francesca Fallucchi, and Suresh Manandhar. 2010. Estimating Linear Models for Compositional Distributional Semantics. In Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010), pages 1263–1271, Beijing, China. Coling 2010 Organizing Committee.

Lei Zhang, Fanchao Qi, Zhiyuan Liu, Yasheng Wang, Qun Liu, and Maosong Sun. 2020. Multi-Channel Reverse Dictionary Model. Proceedings of the AAAI Conference on Artificial Intelligence, 34(01):312–319.