

State of the Art in Intent Detection and Slot Filling for Question Answering System: A Systematic Literature Review

Anis Syafiqah Mat Zailan¹, Noor Hasimah Ibrahim Teo², Nur Atiqah Sia Abdullah³, Mike Joy⁴

School of Computing Sciences-College of Computing, Informatics, and Media,
Universiti Teknologi MARA (UiTM), Melaka Branch, Merlimau, Melaka, Malaysia^{1,2}

School of Computing Sciences-College of Computing, Informatics, and Media,
Universiti Teknologi MARA (UiTM), Shah Alam Branch, Shah Alam, Selangor, Malaysia³

Department of Computer Science, University of Warwick, Coventry, CV4 7AL, United Kingdom⁴

Abstract—A Question Answering System (QAS), also known as a chatbot, is a Natural Language Processing (NLP) application that automatically provides accurate responses to questions posed by humans in natural language. Intent Detection and Classification are crucial elements in NLP, especially in a task-oriented dialogue system. In this paper, we conduct a systematic literature review that will perform a comparative analysis of different techniques or algorithms that are being implemented for intent detection and classification with slot filling. The goals of this paper are to identify the distribution, methodology, techniques or algorithms, and evaluation methods, that can be used to develop and construct a model of intent detection and classification with slot filling. This paper also reviews academic documents that have been published from 2019 to 2023, based on a four-step selection process of identification, screening, eligibility, and inclusion, for the selection process. In order to examine these documents, a systematic review was conducted and four main research questions were answered. The results discuss the methodology that can be used for the implementation of intent detection and classification with slot filling, along with the techniques, algorithms and evaluation methods that are widely used and currently implemented by other researchers.

Keywords—Intent detection; intent classification; slot filling; question answering system

I. INTRODUCTION

A Question Answering System (QAS) is a group of natural language texts or a pre-structured database which are used to automatically provide correct and accurate responses to questions posed by humans in human natural language [1, 2]. A QAS is more capable and more efficient in answering the user query than most search engines such as Google, Yahoo, Live.com, Ask, YouTube, Facebook, and Microsoft Bing [3].

Search engines have a remarkable capability; however, the engines provide lists of related websites and resources including documents that match the user's query [2] without regard to their real intention or what the real question asked is. Hence, instead of the user having to search for the most relevant website or result to their query provided by the search engines, a QAS makes it easier for the user since it just displays the necessary information and results by detecting, recognizing, and classifying the intents of the user via their

human natural language, thus providing them with a corresponding yet accurate response and result directly [2, 4]. Natural Language Processing (NLP) is one of the branches of Artificial Intelligence (AI) that studies human-computer interactions [5]. AI has two subsets, which are Natural Language Understanding (NLU) and Natural Language Generation (NLG). Being able to detect the intent of an utterance has been a keen issue in NLU [6] since it uses syntactic and semantic analysis of the text and speech to determine the meaning of a sentence.

Intent Detection (ID) and Classification (IC) act as critical elements in human language or NLU, especially in a task-oriented dialogue system [1, 6, 7, 8]. In particular, ID is usually related to the keywords of utterance, entities or slots, and intent. ID mainly aims to identify the user's true intent from a given utterance [8] in which, according to [9], the user's true intention sometimes does not settle the meaning of their utterance, and vice versa.

Entities or slots refer to the extraction of the associated arguments of the utterance in which every word in the utterance has its entity or slot. The most usual way of its representation is the IOB representation [10], where the initial "B" in the label indicates the beginning of the slot, the "I" indicates an extension of the "B", and lastly "O" refers to a null label, and it commonly labels the other general words in an utterance. The slots or entities that are constructed are identified by the researchers, hence it is meaningful and easy to be understood by humans.

The intent that is detected by the ID model is another crucial part that defines the context of the text, whereby these intents are the ones that classify and capture the true intention of the user, what they are trying to search for or what they are trying to convey [11] to the application or the system. There are numerous possible intents, depending on the utterance, and for an example of an utterance related to books, some of its intent might be SearchBook, PurchaseBook, BookAuthor, BookGenre, and so on. Many researchers have investigated intent recognition in the English language with very good accuracy. However, to the best of our knowledge, there has been little work done on the Malay language for question answering, with exceptions such as [12].

Within this paper, several methods and methodologies have been identified and proposed for developing ID and IC with a slot-filling model. For instance, [11] has performed several experiments on ID and IC by using various techniques such as Seq2seq, Slot-Gated, Capsule NLU, SF-ID, StackPropagation, SlotRefine, GL-GIN, and Joint-BERT. In addition, a model of ID and IC has been proposed with slot filling of JointBERT and Conditional Random Forest (CRF), called JointIDSF, whereby both XLM-R and PhoBERT were used as the utterance encoder [1]. This paper will contribute to better understanding, and thus developing and implementing, the methods of ID and IC with slot filling in various fields.

In addition, we review the existing literature on the development and implementation of ID and IC with slot filling in this paper in which our main objective is to explore which methods or techniques are the most suitable and can achieve higher accuracy in terms of IC performance. The following are the contributions of our review.

- The previous research that has been published regarding ID and IC with slot filling is investigated.
- The methodologies or approaches proposed by previous research are identified and discussed.
- The techniques or algorithms that can be implemented on ID and IC with slot filling are identified.
- A discussion on the validation process and main results of the previous research is conducted.

Furthermore, many previous research articles have been published and have discussed the researcher's implementation on the topic of ID and IC with slot filling. Papers [1, 8, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37] have discussed and presented their approach or implementation on developing ID and IC with slot filling using various frameworks, methodologies, and techniques. Various frameworks, methodologies, approaches, techniques and algorithms have been implemented by these researchers, whereas for instance, [19, 22, 24, 26, 27, 28, 31, 34, 35, 36] have implemented techniques of Bidirectional Long Short-Term Memory ID (BiLSTM) with Conditional Random Forest (CRF) in their implementation of ID and IC with slot filling.

On the other hand, [8, 15, 17, 20, 24, 26, 28, 31, 33] have implemented integration techniques of Bidirectional Encoder Representations from Transformers (BERT) with various other techniques such as CRF, LSTM, BiLSTM, Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), or Regular Expressions (RE). On top of that, there are also published papers with comprehensive reviews [50-51, 53, 55]. The aim of [50-51] is to conduct a literature review about Intents, Intention Mining, and IC and focuses on the review of algorithms, models, and tools that have been implemented in Intention Mining. A review paper on the ID methods in the human-machine dialogue system seeks to advance the study of multi-intent detection methods based on Recurrent Neural Networks [52] and deep neural networks [53]. This paper primarily analyses compares, and summarizes the deep learning methods used in the research of ID in recent years. It also considers how to apply deep learning models to multi-

intent detection tasks. The third paper [55] introduces the methods of two tasks ranging from the independent model to the joint model. It focuses on joint modelling methods based on deep neural networks and analyzes current problems and future development trends of two sub-tasks.

Therefore, one of the aims of this paper is to focus on conducting a systematic literature review about ID and IC with slot filling in which we are not focusing on the implementation, but we are going deeper into the investigation and discussion of the framework, methodology, and techniques or algorithms that can be implemented in ID and IC with slot filling. Furthermore, what we are reviewing differs from [51, 53, 55], which reviews the topic of intention mining, ID methods in the human-machine dialogue system, and methods of two tasks from the independent model to the joint model, respectively.

This paper is divided as follows: Section II presents the conducted systematic review methodology that consists of the definition of research questions, search phases, inclusion and exclusion criteria, and paper eligibility screening. Section III presents the results of the systematic review consisting of answers to the research questions, and finally, the conclusion of the systematic review and our thoughts on directions for future work are presented in Section IV.

II. RESEARCH METHOD

The systematic review (SR) in this paper was conducted by implementing the preferred reporting items for an SR and meta-analysis (PRISMA) approach as done in [13], in which PRISMA is an evidence-based minimum set of items used to guide the development and structure of SRs and other meta-analysis. According to [14], PRISMA is designed to help researchers to perform literature reviews systematically and transparently. and report how the review was done in a manner which leads to the findings. Thus, by implementing the PRISMA approach in our paper, the reviewing protocol includes three steps, namely definition of the research questions, search phases, and specification of inclusion and exclusion criteria. The specification of these steps is described in the following sections.

A. Definition of Research Question

This SR is organized to encompass the range of research examined by classifying and evaluating previous related articles. To correctly explain the coverage rate of existing works, the research questions must first be defined. We can gain various insights by examining comparable works, which can subsequently assist researchers in coming up with new insights. Table I lists the research questions that were considered in our SR.

B. Search Phase

Defining the information sources is the initial stage in conducting our SR. As shown in Table II, several academic databases, digital libraries, and open-access search engines have been consulted. In order to locate publications that are pertinent to our setting, the following stage entails creating procedures for examining the scientific and technical documentation that these searches produced. The process is built around these two steps: (i) it is necessary to first

determine the search phrases from the earlier research questions in order to create a list of keywords; (ii) it is necessary to create the queries that will be used to locate and gather all connected results, thus the Boolean operators AND/OR will be used to find and gather all related results in accordance to ID and IC with slot filling. About 221 papers were found overall in the first phase with search terms that might be most pertinent in the title. The search terms used for this paper are shown in Table III.

TABLE I. RESEARCH QUESTION

Research question	Motivation
RQ1. What is the distribution per year, domain application, and publisher of the published papers related to ID and IC with slot filling?	The answer to this question allows us to identify the work's domain, when, and where the research studies have been conducted.
RQ2. What is the methodology used to develop ID and IC with slot filling?	The answer to this question illustrates the steps and phases in developing ID and IC with slot filling.
RQ3. What are the techniques or algorithms that can be used to implement ID and IC with slot filling?	The answer to this question helps to identify the most suitable techniques or algorithms that can be adopted into the implementation of ID and IC with slot filling.
RQ4. Which evaluation method was used and what are the main results that have been drawn based on the evaluation method used?	The answer to this question identifies the methods used to evaluate the performance of ID and IC with slot filling and presents the main results or outcomes of the studied works.

TABLE II. SEARCH SOURCES

Source	Type	URL
Science Direct-Elsevier	Digital Library	https://www.sciencedirect.com/
Scopus	Search Engine	https://www.scopus.com/home.uri
IEEE Explore	Digital Library	https://ieeexplore.ieee.org/Xplore/home.jsp
ACM Digital Library	Digital Library	https://dl.acm.org/
Web of Science	Search Engine	https://www.webofknowledge.com/
SpringerLink	Digital Library	https://link.springer.com/
Google Scholar	Search Engine	https://scholar.google.com/

TABLE III. RESEARCH QUERIES

TITLE-ABS-KEY
S1 (intent AND detection OR recognition AND classification)
S2 (intent AND detection AND classification)
S3 (intent AND detection AND recognition AND classification)
S4 (intent AND detection AND slot filling)
S5 (intent AND classification AND slot filling)
S6 (intent AND detection AND classification AND slot filling)

C. Inclusion and Exclusion Criteria

We employed a set of inclusion criteria and exclusion criteria to identify pertinent papers and to narrow search results (see Table IV). Papers that do not address the exclusion criteria are disregarded, and a screening procedure is used to identify papers that are pertinent to our setting. The following three inclusion criteria phases form the basis of the screening procedure:

1) *An abstract-based step*: It using details and keywords from publication abstracts, we eliminate irrelevant results. Articles were kept for additional screening if their abstracts met at least 60% of the inclusion criteria.

2) *Full-text-based step*: We eliminate any results that did not address or make reference to the research terms listed in Table III, i.e., any publications that only cover a small portion of the search terms mentioned in their abstracts.

3) *Quality-analysis-based step*: We perform quality analysis on the remaining results and discard any that do not meet the requirements listed below:

- C1: The paper discusses a comprehensive approach and methodology to ID and IC with slot filling.
- C2: The paper includes the technical implementation of the proposed solution.
- C3: The paper references additional works.
- C4: The paper discusses the outcomes that were found.

TABLE IV. LIST OF IC AND EC

Inclusion Criteria	Exclusion Criteria
Studies published within the period 2019-2023.	Studies that are not written in English.
Studies should meet at least one of the search terms.	Duplicated papers.
Studies should be published/in-press at a journal or conference.	Studies with missing full text.
Studies should provide answers to the research questions.	Papers that are not directly relevant to the ID and IC with slot-filling topics.
The search is performed based on the title, abstract, and full text.	

D. Systematic Search Strategy Procedure

The PRISMA systematic search-strategy procedure includes four core processes which are identification, screening, eligibility, and inclusion were used to choose the pertinent publications for this review. In the identification procedure initially got 221 records (n=221). After the screening procedure, the results were pared down to 149 (n= 149) after duplicates were eliminated. Next subjected to eligibility criteria based on the title and abstract, acquiring 69 (n=69); finally, eligibility criteria based on the complete text allowed us to acquire 25 pertinent studies. To glean the findings reported in the next part, a thorough analysis of these 25 studies was conducted.

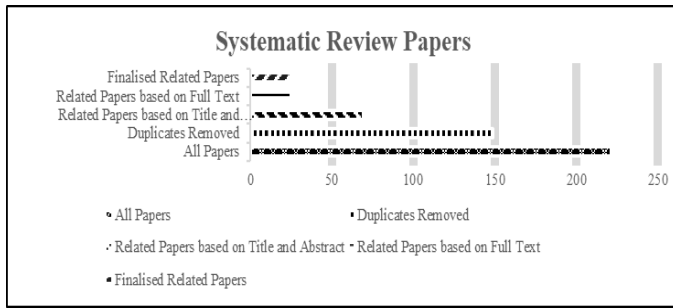


Fig. 1. Related papers reviewed.

Of the 25 papers that had been screened from the review process, all of these papers discussed the ID and IC with slot-filling topics by using various techniques or algorithms. Fig. 1 illustrates the number of papers from each stage in the SR process in a graphical representation of the papers that have been reviewed.

III. RESULTS AND DISCUSSION

This section includes a discussion of the review’s findings about the earlier proposed research questions. This review is made up of 25 publications that were carefully chosen to address the topic of ID and IC with slot filling. These answers help us to know the related recent literature, methodologies used, techniques or algorithms that can be used to implement ID and IC with slot filling, as well as the methods that can be used to evaluate the performance of the model.

Answer to research question RQ1: What is the distribution per year, domain application, and publisher of the published papers related to ID and IC with slot filling?

The papers that have been gathered and which relate to ID and IC with slot filling originated from several domains, including medical, education, electrical, music, economy, and airline. Fig. 2 and Fig. 3 display the distribution of the selected papers by publication year and source respectively.

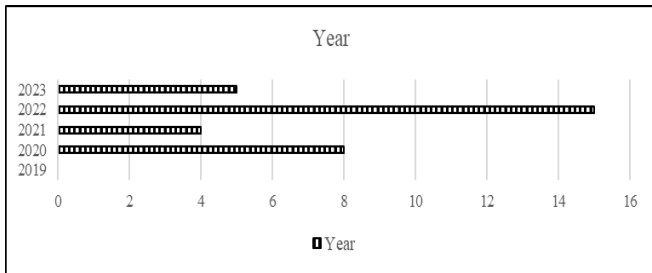


Fig. 2. Distribution of selected papers by publication year.

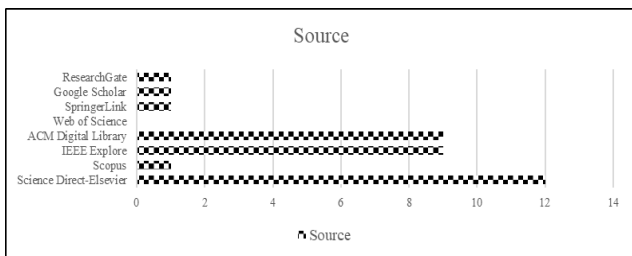


Fig. 3. Distribution of selected papers by source.

Answer to research question RQ2: What is the methodology used to develop ID and IC with slot filling?

Numerous methodologies were used for developing ID and IC with slot-filling phases in which Table V displays the list of methodology names and the phases that are involved in developing ID and IC with slot-filling. After reviewing and analyzing all of the papers, we identified that the majority of the papers integrate several methods resulting in a new methodology or approach. Papers 1, 22, and 25 have integrated Bidirectional Long Short-Term Memory (BiLSTM), Bidirectional Encoder Representations from Transformers (BERT), and Joint BERT with Conditional Random Forest (CRF) respectively. In addition, since there are various and multiple ways in which ID and IC with slot filling have been implemented, almost all of the papers’ implementations differ from one another, even though they might seem similar.

The second paper implements a novel non-aggressive joint model, and the fourth paper implements a GloVe approach in their implementation. Some papers implement a unique methodology or approach such as paper 9 which implements a Capsule Network or Capsule-NLU, paper 13 implements a Multi-level Shared-private Framework, paper 14 implements a Deep Concurrent Multi-Task Paradigm, a Dual pseudo-labelling and dual learning methods approach by Paper 16, a Generative and Classification-based approach is implemented by paper 17, and paper 23 implements an Attention-based RNN and Slot-Gated mechanism. Meanwhile, the remaining methodologies implemented by the remaining papers may be further viewed in Table V and Table VI.

In addition, we removed papers [38-47, 54] and [2-15, 49-50, 56-70] from the list due to their related survey paper status. Furthermore, papers [2-15, 49, 56-70, 81] were removed due to their unrelatedness to ID and IC with slot filling or QAS topic relevance.

Answer to research question RQ3: What are the techniques or algorithms that can be used to implement ID and IC with slot filling?

There are several techniques or algorithms that can be implemented for ID and IC with slot filling. According to Table V, these 25 papers have implemented various techniques. Some papers have similarities in their approach with one another but still differ and report unique results.

Initially, Long Short-Term Memory often referred to as LSTM networks are an extension of Recurrent Neural Network (RNN) whereas RNN is a type of neural network that is specially designed for sequence prediction problems since it imposes an order on the observations that must be preserved when training models and making predictions [72]. The LSTM architecture consists of a set of recurrently connected subnets that are also known as memory blocks, which can be thought of as a differentiable version of memory chips in a digital computer [73]. Not only that, LSTM has been primarily implemented for problems such as speech modelling and language translation [72] and it has achieved state-of-the-art performance for IC and slot filling [74]. In addition, an LSTM network is a special variant of an RNN such that it overcomes stability bottlenecks encountered in traditional RNNs, thus

enabling its practical application [72]. Furthermore, an LSTM can also utilize its internal memory in such a way that its predictions are conditional on the recent context in the input sequence, not what has just been presented as the current input to the network. As an example, an LSTM model can show one observation at a time sequentially, and it can learn what observations it has seen previously and which are relevant. From there, it will think and train on how prediction can be done based on their observations made earlier [72].

Furthermore, Bidirectional Long Short-Term Memory (BiLSTM) is a further refinement of LSTM [24] which integrates the forward hidden layer and the backward hidden layer [71], which can acquire and access both the previous and subsequent contexts. Since LSTM exclusively exploits the historical context, unlike BiLSTM, as a result, BiLSTM is better than LSTM at resolving the sequential modelling problem [71]. LSTM and BiLSTM have been used to classify texts and have achieved some progress [19, 22, 24, 26-28].

According to our review, the majority of the papers implement BiLSTM whereas this technique was implemented by 10 papers - 7, 10, 12, 14, 15, 16, 19, 22, 23, and 24. In addition, these papers have integrated BiLSTM with various other techniques which include Conditional Random Forest (CRF), Concurrent Neural Network (CNN), and Bidirectional Encoder Representations from Transformers (BERT) to perform both ID and IC with slot-filling tasks. Some of these papers only implement BiLSTM such as papers 15 and 23 as their main technique. In addition, papers 7, 22, and 24 have integrated BiLSTM with CRF, papers 10 and 14 have integrated BiLSTM with CNN, and are followed by papers 12 and 16 which have integrated BiLSTM with BERT respectively. However, there is also a paper that integrates BiLSTM with two more techniques, and paper 19 has integrated BiLSTM with both CNN and BERT to perform ID and IC with slot filling. The integration between two to three techniques may help to improve the proposed model's performance.

TABLE V. LIST OF PROPOSED FRAMEWORK AND TECHNIQUES

Reference	Proposed Framework	Technique(s) Used
Paper 1 [1]	JointIDSF	JointBERT+CRF
Paper 2 [8]	SlotRefine	BERT+CRF
Paper 3 [15]	Neural Network-Regular Expressions (NN-RE)	RE+CNN+RNN+ BERT+LSTM
Paper 4 [16]	Computational Linguistics with Deep-Learning-Based Intent Detection and Classification (CL-DLBIDC)	Deep Learning Modified Neural Network (DLMNN) +Mayfly Optimization (MFO)
Paper 5 [17]	SLIM	BERT
Paper 6 [18]	BART+MS+I	BART
Paper 7 [19]	Bi-confidence-frequency cross-lingual transfer framework (BiCF)	BiLSTM+CRF
Paper 8 [20]	MultiLingual MultiTask (MLMT)	BERT+CRF
Paper 9 [21]	Capsule ISNP	NER
Paper 10 [22]	Real-time Pilot-controller Voice Communications (PCVC)	Convolutional Neural Network (CNN) + BiLSTM
Paper 11 [23]	PIL Chatbot with Angular, Flask and RASA framework	JointBERT
Paper 12 [24]	Context-aware Graph Convolutional Network with Adaptive Fusion Layer (CGCN-AF)	BiLSTM+BERT
Paper 13 [25]	No specific framework name mentioned	RoBERTa
Paper 14 [26]	CIDIS - Concurrent Intelligent Model for Dialogue Act Classification, Intent Detection and Slot Filling	BERT+Character Embedding Formulation (CharCNN)+BiLSTM
Paper 15 [27]	No specific framework name mentioned	BiLSTM
Paper 16 [28]	Dual semi-supervised NLU with Semantic-to-sentence Generation (SSG)	BiLSTM+CRF+BERT
Paper 17 [29]	No specific framework name mentioned	JointBERT+XLM-Roberta
Paper 18 [30]	No specific framework name mentioned	KoBERT, KLUE-RoBERTa, mBERT
Paper 19 [31]	Multitask Learning with Knowledge Base for Joint Slot-Filling and Intent-Detection (MTL)	BERT+BiLSTM +CNN
Paper 20 [32]	No specific framework name mentioned	CNN+LSTM+Rules
Paper 21 [33]	Tagger and Classifier	LSTM+BERT
Paper 22 [34]	MTL-Fully Shared Network (MTL-FSN) and Hierarchical-MTL (H-MTL)	BiLSTM+CRF
Paper 23 [35]	Slot-Gated Modeling	BiLSTM
Paper 24 [36]	SF-ID Network	BiLSTM+CRF
Paper 25 [37]	No specific framework name mentioned	JointBERT+CRF

TABLE VI. LIST OF METHODOLOGIES AND THEIR RESULT

Reference	Methodology	Methodology Phases
Paper 1 [1]	Extension of JointBERT + CRF	Encoding Layer, Intermediate Intent-Slot Attention Layer, Decoding Layers of Intent Detection and Slot Filling
Paper 2 [8]	Novel Non-Aggressive Joint Model	BERT Layer, Multi-Head Self Attention Layer, Multi-Head Positional Attention Layer, CRF Layer, Iteration Mechanism, Joint Optimization
Paper 3 [15]	No specific methodology name mentioned	Memory network encoder, Intent determination and slot filling module, Joint optimization
Paper 4 [16]	GloVe approach	Data Preprocessing, Feature Extraction Process using Word Embedding Technique, Intent Detection and Classification Process using Modified Neural Network Model, Parameter Tuning using Mayfly Optimization Algorithm
Paper 5 [17]	Multi-intent Spoken Language Understanding (SLU)	Encoder, Intent Classifier and Slot Classifier, Slot-Intent Classifier
Paper 6 [18]	No specific methodology name mentioned	Intent Classification and Slot Filling, Hate Speech Taxonomy, Counter Speech, BART+MS+I Joint Optimization
Paper 7 [19]	No specific methodology name mentioned	BiCF Mixing, Latent Space Refinement, BiLSTM+CRF Joint Encoder
Paper 8 [20]	CNN and RNN with LSTM and GRU	Embeddings, Dense Layer, Task Specific Layer, BERT+CRF Layer - Slot Filling and Attention
Paper 9 [21]	Capsule Network / Capsule-NLU	POS Tagging, Named Entity Recognition (NER), Word Embeddings
Paper 10 [22]	Deep learning-based	Automatic Speech Recognition (ASR) - Convolutional Neural Network (CNN), Bidirectional Long Short-Term Memory (BiLSTM), Fully Connected (FC) Layer, Controlling Intent Inference (CII), Control Safety Monitoring (CSM)
Paper 11 [23]	Retrieval-based	Conversation Manager, Domain Knowledge Handler, Natural Language Interpreter - Intent Classification And Slot Filling, Joint BERT-Based Model
Paper 12 [24]	No specific methodology name mentioned	Hierarchical Encoder, Context-Aware Graph Convolutional Network, Intent and Dialogue Act Classification, Adaptive Fusion Layer for Slot Filling, Multi-Task Training
Paper 13 [25]	Multi-level Shared-private Framework	Shared-Private Syntactic Encoder, Domain-Aware Sentence-Level Transfer for Intent Detection, Task-Aware Token-Level Transfer for Slot Filling, Joint Training
Paper 14 [26]	Deep concurrent multi-task paradigm	Intelligent Word Embedding Formulation, Contextual Sentence Representation Formulation, Dialogue Act Classification, Intent Detection, Slot Filling, Query Response Retrieval
Paper 15 [27]	Encoder-decoder model	Encoding Layer, Word Embedding, Bi-LSTM Layer, Slot Filling, Intent Detection Layer
Paper 16 [28]	Dual pseudo-labeling and dual learning methods	Sentence Encoding, Intent Classification, Slot Tagging, Semantic-to-Sentence Generation: Encoder, Decoder, Dual Semi-Supervised NLU, Dual Learning Model
Paper 17 [29]	Generative and Classification-based	Encoder, Slot Filling, Intent Detection, Joint Training, Decoder
Paper 18 [30]	No specific methodology name mentioned	Intent Classifier, Slot Classifier, Slot Value Predictor, Value Refiner
Paper 19 [31]	Multitasking Learning	LSTM-CNN Layer, Bi-LSTM + Attention Layer, WordNet Knowledge Base, Joint Optimization, Adaptive Moment Estimation
Paper 20 [32]	Rule-based and Model-based	Text Matching, Encoding, Intent Detection, Slot Filling, Rules Fusing, Joint Optimization
Paper 21 [33]	Weakly-Supervised Dual-Model Learning	Encoder, Word Embedding, Intent Detection-LSTM Layer, Attention Layer, Linear Layer, Joint Optimization
Paper 22 [34]	BiLSTM and CRF Model	Word Embedding, Named Entity Recognition (NER), Semantic Tagging (SemTag), Bi-LSTM+CRF Layer
Paper 23 [35]	Attention-based RNN and Slot-Gated Mechanism	BiLSTM Layer, Intent Attention Layer, Slot Attention Layer, Slot Sequence, Joint Optimization
Paper 24 [36]	Novel Bi-directional Interrelated Model	BiLSTM Layer, Slot Attention, Intent Attention, SF Subnet, SF Iteration Mechanism, ID Subnet, ID Iteration Mechanism, CRF Layer
Paper 25 [37]	JointBERT and CRF Model	BERT Layer, Softmax Layer, WordPiece Tokenizer, CRF Layer, Joint Optimization

In recent times, the BERT framework has been investigated for jointly identifying the intent and slots of an utterance [74, 75]. The model architecture of BERT is a multi-layer bidirectional transformer encoder that is based on the original Transformer model [74, 76-77] whereby it jointly conditions both left and right contexts in the Transformer [75, 77]. Furthermore, the input representation for BERT is a concatenation of WordPiece embeddings [8], positional embeddings, and segment embedding [74].

The BERT model is pre-trained with two strategies on large-scale unlabeled text which refers to Masked Language Model (MLM) and Next Sentence Prediction (NSP) [74-75, 77]. In addition, MLM's function is to randomly mask in order

to avoid a token observing itself in a multi-layered context and on the other hand, NSP aims to capture useful information for sentence pair-oriented tasks [75]. Nonetheless, BERT models greatly contribute to enhancing NLP and it is the most commonly used transformer architecture [78]. Furthermore, the BERT model can be fine-tuned with just one additional output layer to create state-of-the-art models for a wide range of tasks such as in QAS and language inference, without substantial task-specific architecture modifications [77-78,80].

The most implemented technique other than BiLSTM is Bidirectional Encoder Representations from Transformers known as BERT in which BERT has been implemented by nine papers - 2, 3, 5, 8, 12, 14, 16, 19, and 21. These papers

have also integrated BERT with CRF, LSTM, BiLSTM, CNN, RNN, or Regular Expressions (RE). Firstly, paper 5 has implemented BERT as its sole technique for ID and IC with slot filling. However, of the remaining papers, papers 2 and 8 have integrated BERT with CRF, paper 21 has integrated BERT with LSTM, and paper 12 has integrated BERT with BiLSTM. Papers 14 and 19 have integrated BERT with BiLSTM and CNN meanwhile paper 16 has integrated BERT with BiLSTM and CRF. There is also a paper that has integrated BERT with four more techniques, making a total of 5 techniques including BERT. Paper 3 has integrated BERT with LSTM, CNN, RNN, and Regular Expressions (RE).

In addition, BERT can be extended into several extensions such as JointBERT, RoBERTa, KoBERT, mBERT, and KLUE-RoBERTa. Even though these techniques are extensions of BERT, they still differ and are different in terms of performance, whereas JointBERT has been implemented by papers 1, 11, 17, and 25. Paper 11 only implements JointBERT in its implementation, meanwhile, papers 1 and 25 have integrated JointBERT with CRF, and paper 17 has integrated JointBERT with RoBERTa. This is followed by RoBERTa which has been implemented by papers 13 and 17 and finally, paper 18 has integrated KoBERT, KLUE-RoBERTa, and mBERT in their ID and IC with slot filling implementation.

The remaining papers (3, 4, 6, 9, and 20) have implemented a technique that is not vastly and widely implemented such as some of the papers have implemented Concurrent Neural Networks (CNN), Recurrent Neural Network (RNN), Deep Learning Modified Neural Network (DLMNN), Mayfly Optimization (MFO), Named Entity Recognition (NER) in their implementation of ID and IC with slot filling. More details on each of these paper's implementation of techniques or algorithms can be further viewed in Table V.

Answer to research question RQ4: Which evaluation method was used and what are the main results that have been drawn based on the evaluation method used?

There are various ways of evaluating the performance of a model after it has been implemented by using various techniques. However, there are several main evaluation metrics or methods that are generally used to evaluate the performance of an ID and IC with a filling model. Such methods include accuracy, F1-score, precision, and recall which is mainly applied for both ID and slot filling task, respectively. Besides that, there are also other evaluations conducted within the 25 papers such as sentence accuracy, semantic error, G-measure, and so on. Technically, model evaluation is crucial to check and validate whether the model is well-functioning and works correctly according to the specifications and the requirements or not.

The majority of these papers evaluate the performance of their ID and IC with slot filling model by using accuracy and F1-score for both intent, sentence and slot performance.

Initially, accuracy is the most used empirical measure and it can be defined as a ratio of accurately classified data items to the total number of observations [48, 80], making it one of the most suitable methods to evaluate the performance of ID and IC. Despite it being the most used technique for testing, however, accuracy does not distinguish between the number of correct labels of different classes [79] and is valid only when the evaluation of classification is well-balanced and is not skewed, and there is no class imbalance. Hence, it is not the most appropriate performance metric in some situations especially in a case where the target variable classes in the dataset are unbalanced. This is because, when the model predicts that each point belongs to the majority class label, the accuracy will be high but the model is not accurate because of the imbalances.

Next is a precision technique which is a measure of correctness that is achieved in true prediction, or it also means how many predictions are positive out of all the total of positive predictions. Precision is calculated by the ratio of the total number of correctly classified positive classes divided by the total number of predicted positive classes [68, 80]. However, precision can only work properly and accurately when the FP is higher than the FN [48, 68, 79] and is usually suitable for a system that has a yes or no, true or false result such as e-mail spam detection.

The following evaluation technique is a recall technique which is measured as the actual observations which are predicted correctly. Recall can also be called sensitivity and is the most suitable evaluation technique to be used when the researcher wants to capture as many positives as possible [48, 68, 80]. Recall is a ratio of the total number of correctly classified positive classes divided by the total number of positive classes, or in other words, out of the observations that are positive, how many of them have been predicted by the algorithm [48].

On the other hand, the F1-score, also known as the F-measure, uses both precision and recall and is one of the best evaluation techniques to calculate the performance of an algorithm [48, 80], especially in the evaluation of text classification and identification tasks [59, 80], because it balances out the precision and recall, whereby if the precision is low, the F1 is low and if the recall is low, then the F1 is also low. Plus, the F1score evaluation method has been generally used to evaluate the performance of the slot-filling task.

Therefore, in this paper, we have identified the evaluation methods used in these papers, and Table VII depicts the list of evaluation methods and their results.

TABLE VII. LIST OF EVALUATION METHODS AND THEIR RESULTS

Reference	Evaluation Methods	Results			
		Encoder	Intent Accuracy (%)	Slot F1 (%)	Sentence Accuracy (%)
Paper 1 [1]	Accuracy, F1-Score	XLM-R	97.56	94.95	86.17

		PhoBERT	97.62	94.98	86.25																							
Paper 2 [8]	Accuracy	Slot Accuracy: 96.22% Intent Accuracy: 97.11% Sentence Accuracy: 86.96%																										
Paper 3 [15]	Precision, Recall, F1-Score	<table border="1"> <thead> <tr> <th>Dataset</th> <th>Cat</th> <th>Prec(%)</th> <th>Recall (%)</th> <th>F1-Score (%)</th> </tr> </thead> <tbody> <tr> <td rowspan="2">KVRET</td> <td>Intent</td> <td>98.50</td> <td>98.46</td> <td>98.48</td> </tr> <tr> <td>Slot</td> <td>74.32</td> <td>79.18</td> <td>76.91</td> </tr> <tr> <td rowspan="2">Frames</td> <td>Intent</td> <td>95.32</td> <td>92.95</td> <td>94.17</td> </tr> <tr> <td>Slot</td> <td>74.23</td> <td>73.15</td> <td>73.53</td> </tr> </tbody> </table>	Dataset	Cat	Prec(%)	Recall (%)	F1-Score (%)	KVRET	Intent	98.50	98.46	98.48	Slot	74.32	79.18	76.91	Frames	Intent	95.32	92.95	94.17	Slot	74.23	73.15	73.53			
Dataset	Cat	Prec(%)	Recall (%)	F1-Score (%)																								
KVRET	Intent	98.50	98.46	98.48																								
	Slot	74.32	79.18	76.91																								
Frames	Intent	95.32	92.95	94.17																								
	Slot	74.23	73.15	73.53																								
Paper 4 [16]	Accuracy, Recall, Specificity, F1-Score, MCC, G-Measure	<table border="1"> <thead> <tr> <th>Methods</th> <th>Training Set (%)</th> <th>Testing Set (%)</th> </tr> </thead> <tbody> <tr> <td>Accuracy</td> <td>99.29</td> <td>99.51</td> </tr> <tr> <td>Recall</td> <td>97.50</td> <td>98.26</td> </tr> <tr> <td>Specificity</td> <td>99.58</td> <td>99.71</td> </tr> <tr> <td>F1-Score</td> <td>97.50</td> <td>98.29</td> </tr> <tr> <td>MCC</td> <td>97.09</td> <td>98.01</td> </tr> <tr> <td>G-Measure</td> <td>97.51</td> <td>98.30</td> </tr> </tbody> </table>	Methods	Training Set (%)	Testing Set (%)	Accuracy	99.29	99.51	Recall	97.50	98.26	Specificity	99.58	99.71	F1-Score	97.50	98.29	MCC	97.09	98.01	G-Measure	97.51	98.30					
Methods	Training Set (%)	Testing Set (%)																										
Accuracy	99.29	99.51																										
Recall	97.50	98.26																										
Specificity	99.58	99.71																										
F1-Score	97.50	98.29																										
MCC	97.09	98.01																										
G-Measure	97.51	98.30																										
Paper 5 [17]	F1-Score, Accuracy	<table border="1"> <thead> <tr> <th>Dataset</th> <th>Slot F1 (%)</th> <th>Intent Accuracy (%)</th> <th>SeFr Accuracy (%)</th> </tr> </thead> <tbody> <tr> <td>MixATIS</td> <td>88.5</td> <td>78.30</td> <td>47.60</td> </tr> <tr> <td>MixSNIPS</td> <td>96.5</td> <td>97.20</td> <td>84.00</td> </tr> </tbody> </table>	Dataset	Slot F1 (%)	Intent Accuracy (%)	SeFr Accuracy (%)	MixATIS	88.5	78.30	47.60	MixSNIPS	96.5	97.20	84.00														
Dataset	Slot F1 (%)	Intent Accuracy (%)	SeFr Accuracy (%)																									
MixATIS	88.5	78.30	47.60																									
MixSNIPS	96.5	97.20	84.00																									
Paper 6 [18]	F1-Score	F1-Score on Full Parse Tree: 52.96% F1-Score on Top Level Parse Tree: 56.29% F1-Score on Lower Level Parse Tree: 62.04% F1-Score on Intent Classification: 57.17%																										
Paper 7 [19]	Accuracy	<table border="1"> <thead> <tr> <th>Domain</th> <th>Intent Accuracy (%)</th> <th>Slots Accuracy (%)</th> </tr> </thead> <tbody> <tr> <td>Restaurant</td> <td>93.02</td> <td>82.91</td> </tr> <tr> <td>Hotel</td> <td>94.73</td> <td>77.15</td> </tr> <tr> <td>Taxi</td> <td>92.73</td> <td>91.03</td> </tr> <tr> <td>Attraction</td> <td>94.88</td> <td>90.74</td> </tr> </tbody> </table>	Domain	Intent Accuracy (%)	Slots Accuracy (%)	Restaurant	93.02	82.91	Hotel	94.73	77.15	Taxi	92.73	91.03	Attraction	94.88	90.74											
Domain	Intent Accuracy (%)	Slots Accuracy (%)																										
Restaurant	93.02	82.91																										
Hotel	94.73	77.15																										
Taxi	92.73	91.03																										
Attraction	94.88	90.74																										
Paper 8 [20]	Accuracy, F1-Score	<table border="1"> <thead> <tr> <th>Dataset</th> <th>Intent Accuracy (%)</th> <th>Slots F1-Score (%)</th> <th>Template Accuracy</th> </tr> </thead> <tbody> <tr> <td>ATIS</td> <td>99.18</td> <td>97.93</td> <td>90.05</td> </tr> <tr> <td>Trains</td> <td>86.45</td> <td>99.01</td> <td>89.56</td> </tr> <tr> <td>Frames</td> <td>80.91</td> <td>91.67</td> <td>83.56</td> </tr> <tr> <td>Snips</td> <td>99.11</td> <td>97.08</td> <td>91.20</td> </tr> </tbody> </table>	Dataset	Intent Accuracy (%)	Slots F1-Score (%)	Template Accuracy	ATIS	99.18	97.93	90.05	Trains	86.45	99.01	89.56	Frames	80.91	91.67	83.56	Snips	99.11	97.08	91.20						
Dataset	Intent Accuracy (%)	Slots F1-Score (%)	Template Accuracy																									
ATIS	99.18	97.93	90.05																									
Trains	86.45	99.01	89.56																									
Frames	80.91	91.67	83.56																									
Snips	99.11	97.08	91.20																									
Paper 9 [21]	Accuracy, F1-Score, Semantic Error	<table border="1"> <thead> <tr> <th>Dataset</th> <th>Intent Accuracy (%)</th> <th>Slots F1-Score (%)</th> <th>Semantic Error (%)</th> </tr> </thead> <tbody> <tr> <td>ATIS</td> <td>89.00</td> <td>94.40</td> <td>78.10</td> </tr> <tr> <td>Snips</td> <td>98.00</td> <td>92.90</td> <td>85.00</td> </tr> </tbody> </table>	Dataset	Intent Accuracy (%)	Slots F1-Score (%)	Semantic Error (%)	ATIS	89.00	94.40	78.10	Snips	98.00	92.90	85.00														
Dataset	Intent Accuracy (%)	Slots F1-Score (%)	Semantic Error (%)																									
ATIS	89.00	94.40	78.10																									
Snips	98.00	92.90	85.00																									
Paper 10 [22]	Precision, F1-Score	- Precision: 99.4% - F1-Score: 98.7%																										
Paper 11 [23]	Accuracy, SUS score, Understanding score, Navigation score, Intelligence score	- Accuracy: 80.06% - SUS score: 85.42% - Understanding score: 83.59% - Navigation score: 79.93% - Intelligence score: 92.61%																										

Paper 12 [24]	Accuracy, F1-Score	- Intent Accuracy: 99.96% - Act F1-Score: 97.67% - Slot F1-Score: 95.06% - Frame Accuracy: 91.10%																
Paper 13 [25]	Overall Exact, Accuracy, Exact	<table border="1"> <thead> <tr> <th>Dataset Evaluation</th> <th>MTOD</th> <th>ASMixed</th> </tr> </thead> <tbody> <tr> <td>Overall Exact (%)</td> <td>91.27</td> <td>84.81</td> </tr> <tr> <td>Slot Accuracy (%)</td> <td>95.69</td> <td>94.30</td> </tr> <tr> <td>Intent Accuracy (%)</td> <td>99.20</td> <td>97.30</td> </tr> <tr> <td>Exact (%)</td> <td>Reminder: 85.62 Alarm: 92.37 Weather: 93.29</td> <td>ATIS: 86.53 Snips: 82.62</td> </tr> </tbody> </table>	Dataset Evaluation	MTOD	ASMixed	Overall Exact (%)	91.27	84.81	Slot Accuracy (%)	95.69	94.30	Intent Accuracy (%)	99.20	97.30	Exact (%)	Reminder: 85.62 Alarm: 92.37 Weather: 93.29	ATIS: 86.53 Snips: 82.62	
		Dataset Evaluation	MTOD	ASMixed														
		Overall Exact (%)	91.27	84.81														
		Slot Accuracy (%)	95.69	94.30														
Intent Accuracy (%)	99.20	97.30																
Exact (%)	Reminder: 85.62 Alarm: 92.37 Weather: 93.29	ATIS: 86.53 Snips: 82.62																
Paper 14 [26]	Accuracy, F1-Score	<table border="1"> <thead> <tr> <th>Dataset</th> <th>Intent Accuracy (%)</th> <th>Slots F1-Score (%)</th> <th>Dialogue Act Accuracy (%)</th> </tr> </thead> <tbody> <tr> <td>ATIS</td> <td>98.54</td> <td>98.89</td> <td>99.10</td> </tr> <tr> <td>Frames</td> <td>63.09</td> <td>93.52</td> <td>48.77</td> </tr> <tr> <td>Trains</td> <td>81.21</td> <td>94.07</td> <td>78.89</td> </tr> </tbody> </table>	Dataset	Intent Accuracy (%)	Slots F1-Score (%)	Dialogue Act Accuracy (%)	ATIS	98.54	98.89	99.10	Frames	63.09	93.52	48.77	Trains	81.21	94.07	78.89
		Dataset	Intent Accuracy (%)	Slots F1-Score (%)	Dialogue Act Accuracy (%)													
		ATIS	98.54	98.89	99.10													
Frames	63.09	93.52	48.77															
Trains	81.21	94.07	78.89															
Paper 15 [27]	Accuracy, F1-Score	Intent Accuracy: 97.24% F1-Score: 98.01%																
Paper 16 [28]	Accuracy, F1-Score	<table border="1"> <thead> <tr> <th>Dataset</th> <th>Intent Accuracy (%)</th> <th>Slots F1-Score (%)</th> </tr> </thead> <tbody> <tr> <td>ATIS</td> <td>99.10</td> <td>96.00</td> </tr> <tr> <td>Snips</td> <td>99.10</td> <td>97.10</td> </tr> </tbody> </table>	Dataset	Intent Accuracy (%)	Slots F1-Score (%)	ATIS	99.10	96.00	Snips	99.10	97.10							
		Dataset	Intent Accuracy (%)	Slots F1-Score (%)														
ATIS	99.10	96.00																
Snips	99.10	97.10																
Paper 17 [29]	Accuracy, F1-Score	Intent Accuracy: 96.26% Slot F1-Score: 94.01%																
Paper 18 [30]	Accuracy, F1-Score	<table border="1"> <thead> <tr> <th>Dataset</th> <th>Intent Accuracy (%)</th> <th>Slots F1-Score (%)</th> </tr> </thead> <tbody> <tr> <td>KoBERT</td> <td>98.90</td> <td>99.70</td> </tr> <tr> <td>KLUE-RoBERTa</td> <td>98.98</td> <td>99.45</td> </tr> <tr> <td>mBERT</td> <td>98.38</td> <td>99.52</td> </tr> </tbody> </table>	Dataset	Intent Accuracy (%)	Slots F1-Score (%)	KoBERT	98.90	99.70	KLUE-RoBERTa	98.98	99.45	mBERT	98.38	99.52				
		Dataset	Intent Accuracy (%)	Slots F1-Score (%)														
		KoBERT	98.90	99.70														
KLUE-RoBERTa	98.98	99.45																
mBERT	98.38	99.52																
Paper 19 [31]	Accuracy, F1-Score	<table border="1"> <thead> <tr> <th>Dataset</th> <th>Intent Accuracy (%)</th> <th>Slots F1-Score (%)</th> </tr> </thead> <tbody> <tr> <td>ATIS</td> <td>98.83</td> <td>97.06</td> </tr> <tr> <td>Snips</td> <td>98.79</td> <td>97.31</td> </tr> </tbody> </table>	Dataset	Intent Accuracy (%)	Slots F1-Score (%)	ATIS	98.83	97.06	Snips	98.79	97.31							
		Dataset	Intent Accuracy (%)	Slots F1-Score (%)														
ATIS	98.83	97.06																
Snips	98.79	97.31																
Paper 20 [32]	Accuracy, F1-Score	Accuracy: 97.37% F1-Score: 86.67%																
Paper 21 [33]	Precision, Recall, F1-Score	Precision: 90.5% Recall: 92.40% F1-Score: 91.40%																
Paper 22 [34]	F1-Score	F1-Score for NER: 88.76% F1-Score for Semantic Tagging(SemTag): 88.96% F1-Score for both NER and SemTag: 88.78%																
Paper 23 [35]	Accuracy, F1-Score	<table border="1"> <thead> <tr> <th>Dataset</th> <th>Intent Accuracy (%)</th> <th>Slots F1-Score (%)</th> <th>Sentence Accuracy (%)</th> </tr> </thead> <tbody> <tr> <td>ATIS</td> <td>94.10</td> <td>95.20</td> <td>82.60</td> </tr> <tr> <td>Snips</td> <td>96.80</td> <td>88.30</td> <td>74.60</td> </tr> </tbody> </table>	Dataset	Intent Accuracy (%)	Slots F1-Score (%)	Sentence Accuracy (%)	ATIS	94.10	95.20	82.60	Snips	96.80	88.30	74.60				
		Dataset	Intent Accuracy (%)	Slots F1-Score (%)	Sentence Accuracy (%)													
ATIS	94.10	95.20	82.60															
Snips	96.80	88.30	74.60															
Paper 24 [36]	Accuracy, F1-Score	<table border="1"> <thead> <tr> <th>Dataset</th> <th>Intent Accuracy (%)</th> <th>Slots F1-Score (%)</th> <th>Sentence Accuracy (%)</th> </tr> </thead> <tbody> <tr> <td>ATIS</td> <td>97.09</td> <td>95.80</td> <td>86.90</td> </tr> <tr> <td>Snips</td> <td>97.29</td> <td>92.23</td> <td>80.43</td> </tr> </tbody> </table>	Dataset	Intent Accuracy (%)	Slots F1-Score (%)	Sentence Accuracy (%)	ATIS	97.09	95.80	86.90	Snips	97.29	92.23	80.43				
		Dataset	Intent Accuracy (%)	Slots F1-Score (%)	Sentence Accuracy (%)													
ATIS	97.09	95.80	86.90															
Snips	97.29	92.23	80.43															

Paper 25 [37]	Accuracy, F1-Score	Dataset	Intent Accuracy (%)	Slots F1-Score (%)	Sentence Accuracy (%)
		ATIS	98.40	96.70	92.30
		Snips	97.90	96.00	88.60

TABLE VIII. LIST OF DATASET, DOMAIN AND LANGUAGE

Reference	Dataset	Domain	Language
Paper 1 [1]	ATIS	Airline	Vietnamese
Paper 2 [8]	ATIS, SNIPS	Airline, General	English
Paper 3 [15]	KVRET, FRAMES	In-Car Assistant, Hotel and Travel-Booking	English
Paper 4 [16]	SNIPS	General	English
Paper 5 [17]	ATIS, SNIPS	Airline, General	English
Paper 6 [18]	Policy-aware Explainable Abuse Detection (PLEAD)	Human Abuse	English
Paper 7 [19]	Self-established ID-WOZ	Restaurant, Hotel, Taxi, Attraction	Indonesian
Paper 8 [20]	ATIS, TRAINS, SNIPS, and FRAMES	Airline, Trains, General, In-Car Assistant, Hotel and Travel-Booking	Hindi, Bengali
Paper 9 [21]	ATIS, SNIPS	Airline, General	English
Paper 10 [22]	Self-Collected Pilot-Controller Voice Communications (PCVC)	Airline	Chinese, English
Paper 11 [23]	PILs Model Repository	Drugs / Medicine	Italian
Paper 12 [24]	ATIS	Airline	English
Paper 13 [25]	MTOD, ASMixed (ATIS, SNIPS)	Alarm, Reminder, Weather, Airline, General	English
Paper 14 [26]	ATIS, TRAINS, FRAMES	Airline, Trains, Hotel and Travel-Booking	English
Paper 15 [27]	ATIS, DSTC5 (The Fifth Dialog State Tracking Challenge)	Airline, State	English
Paper 16 [28]	ATIS, SNIPS	Airline, General	English
Paper 17 [29]	ATIS	Airline	Tamil
Paper 18 [30]	In-Vehicle Domain Dialogue Data	Vehicle	Korean
Paper 19 [31]	ATIS, SNIPS, FRAMES, TRAINS	Airline, General, Hotel and Travel-Booking, Trains	English, Hindi, Bengali
Paper 20 [32]	Self-Collected Music and Non-Music Field of the Human-Machine Dialogue System	Music	English
Paper 21 [33]	ATIS, SNIPS, MIT Restaurant	Airline, General, Restaurant	English
Paper 22 [34]	ATIS, MIT Restaurant, MIT Movie	Airline, Restaurant, Movie	English
Paper 23 [35]	ATIS, SNIPS	Airline, General	English
Paper 24 [36]	ATIS, SNIPS	Airline, General	English
Paper 25 [37]	ATIS, SNIPS	Airline, General	English

In addition, each work's dataset, domain, and language used for the implementation are also written in Table VIII whereby most of the works are developed in English. Therefore, few papers have developed this topic on low-resource languages other than English, such as Bengali [20, 31], Chinese [22], Hindi [20, 31], Indonesian [19], Italian [23], Korean [30], Tamil [29] and Vietnamese [1].

IV. CONCLUSION

This paper summarizes the distribution of papers per year, domain application, and source of the published papers related to ID and IC with slot filling along with the proposed frameworks, techniques or algorithms used, the methodology that has been implemented research, as well as the methodology phases, in each paper. In addition, a systematic review has been conducted using the PRISMA approach, and its selection process of identification, screening, eligibility, and inclusion was reported in detail. A total of 25 works were selected from the 221 works that have been initially extracted, based on their relevance to the four main research questions we have developed. In addition, from the review, we discovered that the techniques and algorithms that are generally and widely used to implement ID and IC with slot filling are Bidirectional Long Short-Term Memory (BiLSTM), Bidirectional Encoder Representations from Transformers (BERT), and Conditional Random Forest (CRF). Not only that, for the evaluation methods, we have looked at various evaluation techniques for ID and IC with slot filling, and we have identified that the majority of the past works' models have been evaluated by the accuracy and F1-score evaluation methods. Therefore, our review on this topic has led us to conclude that ID and IC with slot filling are still crucial and indeed still in need of evolution especially for the low-resource languages other than English such as Malay, Chinese, Tamil, or Vietnamese. Hence, the development of ID and IC with slot filling for low-resource languages requires further studies, implementation, and optimization, in order to provide timely future work opportunities for researchers who are interested in this integrative field.

In the future, this research on intent and slot-filling recognition aims to make these systems more accurate, adaptable to different domains, and responsive in real-time interactions. The focus will be on combining various data sources, like text, speech, and images, to better understand user intents, making conversations more natural. Personalizing models for individual user preferences and ensuring ethical considerations, such as minimizing biases, will be crucial. Additionally, efforts will be directed toward making models interpretable and capable of handling multiple languages seamlessly.

ACKNOWLEDGMENT

The authors express their sincere appreciation to the reviewers for their valuable feedback. The funding for this research was provided by Universiti Teknologi MARA under the Young Talent Research Grant (600-RMC/YTR/5/3 (023/2020)).

REFERENCES

- [1] Dao, M. H., Truong, T. H., & Nguyen, D. Q. (2021). Intent detection and slot filling for Vietnamese. Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH, 5, 3916–3920. <https://doi.org/10.21437/Interspeech.2021-618>.
- [2] Sweta P. Lende, M. M. R. (2021). Closed domain question answering system using NLP techniques. January 2016. <https://doi.org/10.5281/zenodo.49808>.
- [3] Murphy E P, Fenelon C, Murphy F, et al. (2019). Does Google Have the Answers? The Internet-based Information on Pelvic and Acetabular Fractures. *Cureus* 11(10): e5952. <https://doi.org/10.7759/cureus.5952>.
- [4] Tang, Y., Han, H., Yu, X., Zhao, J., Liu, G., & Wei, L. (2021). An Intelligent Question Answering System based on the Power Knowledge Graph. 1–5. <https://doi.org/https://doi.org/10.48550/arXiv.2106.09013>
- [5] Dowlagar, S., & Mamidi, R. (2023). A code-mixed task-oriented dialogue dataset for the medical domain. *Computer Speech and Language*, 78(September 2022), 101449. <https://doi.org/10.1016/j.csl.2022.101449>.
- [6] Zhang, C., Li, Y., Du, N., Fan, W., & Yu, P. S. (2020). Joint slot filling and intent detection via capsule neural networks. *ACL 2019 - 57th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference*, 5259–5267. <https://doi.org/10.18653/v1/p19-1519>.
- [7] Alshahrani, H. J., Tarmissi, K., Alshahrani, H., Ahmed Elfaki, M., Yafoz, A., Alsini, R., Alghushairy, O., & Ahmed Hamza, M. (2022). Computational Linguistics with Deep-Learning-Based Intent Detection for Natural Language Understanding. *Applied Sciences (Switzerland)*, 12(17). <https://doi.org/10.3390/app12178633>.
- [8] Wu, D., Ding, L., Lu, F., & Xie, J. (2020). SlotRefine: A fast non-autoregressive model for joint intent detection and slot filling. *EMNLP 2020 - 2020 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference*, 1932–1937. <https://doi.org/10.18653/v1/2020.emnlp-main.152>.
- [9] Leth, P. (2021). Utterance Interpretation and Actual Intentions. *Axiomathes*, 31(3), 279–298. <https://doi.org/10.1007/s10516-019-09462-x>.
- [10] D. Guo, G. Tur, W. -t. Yih and G. Zweig, "Joint semantic utterance classification and slot filling with recursive neural networks," 2014 IEEE Spoken Language Technology Workshop (SLT), South Lake Tahoe, NV, USA, 2014, pp. 554-559, doi: 10.1109/SLT.2014.7078634.
- [11] Gangadharaiyah, R., & Narayanaswamy, B. (2019). Joint Multiple Intent Detection and Slot Labeling for Goal-Oriented Dialog. *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*, 1, 564–569. <https://doi.org/10.18653/v1/n19-1055>.
- [12] Puteh, N., Husin, M. Z., Tahir, H. M., & Hussain, A. (2019). Building a Question Classification Model for a Malay Question Answering System. *International Journal of Innovative Technology and Exploring Engineering*, 8(5s), 184–190.
- [13] Zulkipli, Z. Z., Maskat, R., & Teo, N. H. I. (2022). A systematic literature review of automatic ontology construction. *Indonesian Journal of Electrical Engineering and Computer Science*, 28(2), 878–889. <https://doi.org/10.11591/ijeecs.v28.i2.pp878-889>.
- [14] M. J. Page et al., "The PRISMA 2020 statement: an updated guideline for reporting systematic reviews," *BMJ*, vol. 372, p. n71, Mar. 2021, doi: 10.1136/bmj.n71.
- [15] Abro, W. A., Qi, G., Ali, Z., Feng, Y., & Aamir, M. (2020). Multi-turn intent determination and slot filling with neural networks and regular expressions. *Knowledge-Based Systems*, 208, 106428. <https://doi.org/10.1016/j.knsys.2020.106428>.
- [16] Alshahrani, H. J., Tarmissi, K., Alshahrani, H., Ahmed Elfaki, M., Yafoz, A., Alsini, R., Alghushairy, O., & Ahmed Hamza, M. (2022). Computational Linguistics with Deep-Learning-Based Intent Detection for Natural Language Understanding. *Applied Sciences (Switzerland)*, 12(17). <https://doi.org/10.3390/app12178633>.

- [17] Cai, F., Zhou, W., Mi, F., & Faltings, B. (2022). Slim: Explicit Slot-Intent Mapping with Bert for Joint Multi-Intent Detection and Slot Filling. ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings, 2022-May, 7607–7611. <https://doi.org/10.1109/ICASSP43922.2022.9747477>.
- [18] Calabrese, A., Ross, B., & Lapata, M. (2022). Explainable Abuse Detection as Intent Classification and Slot Filling. Transactions of the Association for Computational Linguistics, 10, 1440–1454. https://doi.org/10.1162/tacl_a_00527.
- [19] Di, D., Song, X., Zhang, W., Zhang, Y., & Wang, F. (2022). Building Dialogue Understanding Models for Low-resource Language Indonesian from Scratch. ACM Transactions on Asian and Low-Resource Language Information Processing. <https://doi.org/10.1145/3575803>.
- [20] Firdaus, M., Ekbal, A., & Cambria, E. (2023). Multitask learning for multilingual intent detection and slot filling in dialogue systems. Information Fusion, 91(October 2022), 299–315. <https://doi.org/10.1016/j.inffus.2022.09.029>.
- [21] Staliunaite, I., & Iacobacci, I. (2020). Auxiliary Capsules for Natural Language Understanding. ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 8149–8153. <https://doi.org/https://doi.org/10.1109/ICASSP40776.2020.9053899>.
- [22] Lin, Y., Deng, L., Chen, Z., Wu, X., Zhang, J., & Yang, B. (2020). A Real-Time ATC Safety Monitoring Framework Using a Deep Learning Approach. IEEE Transactions on Intelligent Transportation Systems, 21(11), 4572–4581. <https://doi.org/10.1109/TITS.2019.2940992>.
- [23] Minutolo, A., Damiano, E., De Pietro, G., Fujita, H., & Esposito, M. (2022). A conversational agent for querying Italian Patient Information Leaflets and improving health literacy. Computers in Biology and Medicine, 141(August 2021), 105004. <https://doi.org/10.1016/j.combiomed.2021.105004>.
- [24] Qin, L., Che, W., Ni, M., Li, Y., & Liu, T. (2021). Knowing Where to Leverage: Context-Aware Graph Convolutional Network with an Adaptive Fusion Layer for Contextual Spoken Language Understanding. IEEE/ACM Transactions on Audio Speech and Language Processing, 29, 1280–1289. <https://doi.org/10.1109/TASLP.2021.3053400>.
- [25] Qin, L., Wei, F., Ni, M., Zhang, Y., Che, W., Li, Y., & Liu, T. (2022). Multi-domain Spoken Language Understanding Using Domain-And Task-Aware Parameterization. ACM Transactions on Asian and Low-Resource Language Information Processing, 21(4), 1–17. <https://doi.org/10.1145/3502198>.
- [26] Sunilkumar, G., Srihari, S., Gilbert, S. F., & Chitrakala, S. (2022). A Concurrent Intelligent Natural Language Understanding Model for an Automated Inquiry System. Proceedings - 2022 IEEE World Conference on Applied Intelligence and Computing, AIC 2022, 124–129. <https://doi.org/10.1109/AIC55036.2022.9848883>.
- [27] Xu, C., Li, Q., Zhang, D., Cui, J., Sun, Z., & Zhou, H. (2020). A model with length-variable attention for spoken language understanding. Neurocomputing, 379, 197–202. <https://doi.org/10.1016/j.neucom.2019.07.112>.
- [28] Zhu, S., Cao, R., & Yu, K. (2020). Dual Learning for Semi-Supervised Natural Language Understanding. IEEE/ACM Transactions on Audio Speech and Language Processing, 28, 1936–1947. <https://doi.org/10.1109/TASLP.2020.3001684>.
- [29] Ramaneswaran, S., Vijay, S., & Srinivasan, K. (2022). TamilATIS: Dataset for Task-Oriented Dialog in Tamil. Dravidian-LangTech 2022 - 2nd Workshop on Speech and Language Technologies for Dravidian Languages, Proceedings of the Workshop, 25–32. <https://doi.org/10.18653/v1/2022.dravidianlangtech-1.4>.
- [30] Lim, J., Son, S., Lee, S., Chun, C., Park, S., Hur, Y., & Lim, H. (2022). Intent Classification and Slot Filling Model for In-Vehicle Services in Korea. Applied Sciences (Switzerland), 12(23). <https://doi.org/10.3390/app122312438>.
- [31] He, T., Xu, X., Wu, Y., Wang, H., & Chen, J. (2021). Multitask learning with a knowledge base for joint intent detection and slot filling. Applied Sciences (Switzerland), 11(11). <https://doi.org/10.3390/app11114887>.
- [32] Ren, S., Wang, H., Yu, D., Li, Y., & Li, Z. (2018). Joint intent detection and slot filling with rules. CEUR Workshop Proceedings, 2242, 34–40.
- [33] Wang, J., Chen, K., Shou, L., Wu, S., & Chen, G. (2021). Effective Slot Filling via Weakly-Supervised Dual-Model Learning. 35th AAAI Conference on Artificial Intelligence, AAAI 2021, 16, 13952–13960. <https://doi.org/10.1609/aaai.v35i16.17643>.
- [34] Louvan, S., & Magnini, B. (2019). Leveraging non-conversational tasks for low resource slot filling: Does it help? SIGDIAL 2019 - 20th Annual Meeting of the Special Interest Group Discourse Dialogue - Proceedings of the Conference, 2003(September), 85–91. <https://doi.org/10.18653/v1/w19-5911>.
- [35] Goo, C. W., Gao, G., Hsu, Y. K., Huo, C. L., Chen, T. C., Hsu, K. W., & Chen, Y. N. (2018). Slot-gated modelling for joint slot filling and intent prediction. NAACL HLT 2018 - 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference, 2, 753–757. <https://doi.org/10.18653/v1/n18-2118>.
- [36] Haihong, E., Niu, P., Chen, Z., & Song, M. (2020). A novel bi-directional interrelated model for joint intent detection and slot filling. ACL 2019 - 57th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference, September, 5467–5471. <https://doi.org/10.18653/v1/p19-1544>.
- [37] Chen, Q., Zhuo, Z., & Wang, W. (2019). BERT for Joint Intent Classification and Slot Filling. 1–11. <http://arxiv.org/abs/1902.10909>.
- [38] Deriu, J., Rodrigo, A., Otegi, A., Echegoyen, G., Rosset, S., Agirre, E., & Cieliebak, M. (2021). Survey on evaluation methods for dialogue systems. In Artificial Intelligence Review (Vol. 54, Issue 1). Springer Netherlands. <https://doi.org/10.1007/s10462-020-09866-x>.
- [39] Jannach, D., Manzoor, A., Cai, W., & Chen, L. (2021). A Survey on Conversational Recommender Systems. ACM Computing Surveys, 54(5). <https://doi.org/10.1145/3453154>.
- [40] Keyvan, K., & Huang, J. X. (2022). How to Approach Ambiguous Queries in Conversational Search: A Survey of Techniques, Approaches, Tools, and Challenges. ACM Computing Surveys, 55(6). <https://doi.org/10.1145/3534965>.
- [41] Han, X., Wang, Y. T., Feng, J. L., Deng, C., Chen, Z. H., Huang, Y. A., Su, H., Hu, L., & Hu, P. W. (2023). A survey of trans-former-based multimodal pre-trained models. Neurocomputing, 515, 89–106. <https://doi.org/10.1016/j.neucom.2022.09.136>.
- [42] Gupta, M., & Agrawal, P. (2022). Compression of Deep Learning Models for Text: A Survey. ACM Transactions on Knowledge Discovery from Data, 16(4), 1–55. <https://doi.org/10.1145/3487045>.
- [43] Liu, S., Mallol-Ragolta, A., Parada-Cabaleiro, E., Qian, K., Jing, X., Kathan, A., Hu, B., & Schuller, B. W. (2022). Audio self-supervised learning: A survey. Patterns, 3(12). <https://doi.org/10.1016/j.patter.2022.100616>.
- [44] Fu, T., Gao, S., Zhao, X., Wen, J. Rong, & Yan, R. (2022). Learning towards conversational AI: A survey. AI Open, 3(February), 14–28. <https://doi.org/10.1016/j.aiopen.2022.02.001>.
- [45] Uc-Cetina, V., Navarro-Guerrero, N., Martin-Gonzalez, A., Weber, C., & Wermter, S. (2022). Survey on reinforcement learning for language processing. In Artificial Intelligence Review (Vol. 56, Issue 2). Springer Netherlands. <https://doi.org/10.1007/s10462-022-10205-5>.
- [46] Liu, K., Chen, Y., Liu, J., Zuo, X., & Zhao, J. (2020). Extracting Events and Their Relations from Texts: A Survey on Recent Research Progress and Challenges. AI Open, 1(February), 22–39. <https://doi.org/10.1016/j.aiopen.2021.02.004>.
- [47] Weld, H., Huang, X., Long, S., Poon, J., & Han, S. C. (2023). A Survey of Joint Intent Detection and Slot Filling Models in Natural Language Understanding. ACM Computing Surveys, 55(8), 1–38. <https://doi.org/10.1145/3547138>.
- [48] Vakili, M., Ghamsari, M., & Rezaei, M. (2020). Performance Analysis and Comparison of Machine and Deep Learning Algorithms for IoT Data Classification. <https://doi.org/https://doi.org/10.48550/arXiv.2001.09636>.
- [49] Li, G. (2022). Question Answering System Based on Knowledge Graph in Traditional Chinese Medicine Diagnosis and Treatment of Viral Hepatitis B. 2022.
- [50] Weld, H. (2021). A survey of joint intent detection and slot-filling models in natural language understanding. ACM Transactions on Graphics, 37(4). <https://doi.org/https://doi.org/10.1145/3547138>.

- [51] Yanli, H. (2021). Research on Spoken Language Understanding Based on Deep Learning. Scientific Programming, 2021. <https://doi.org/10.1155/2021/8900304>.
- [52] Liu, Z., Feng, Y., & Chen, Z. (2021). DialTest: Automated testing for recurrent neural network-driven dialogue systems. ISSTA 2021 - Proceedings of the 30th ACM SIGSOFT International Symposium on Software Testing and Analysis, 115–126. <https://doi.org/10.1145/3460319.3464829>.
- [53] Liu, J., Li, Y., & Lin, M. (2019). Review of Intent Detection Methods in the Human-Machine Dialogue System. Journal of Physics: Conference Series, 1267(1). <https://doi.org/10.1088/1742-6596/1267/1/012059>.
- [54] Larson, S., & Leach, K. (2022). A Survey of Intent Classification and Slot-Filling Datasets for Task-Oriented Dialog. 1(Section 5), 1–30. <https://doi.org/https://doi.org/10.48550/arXiv.2207.13211>.
- [55] Hou, L., Li, Y., Li, C., & Lin, M. (2019). Review of Research on Task-Oriented Spoken Language Understanding. Journal of Physics: Conference Series, 1267(1). <https://doi.org/10.1088/1742-6596/1267/1/012023>.
- [56] Dowlagar, S., & Mamidi, R. (2023). A code-mixed task-oriented dialog dataset for the medical domain. Computer Speech and Language, 78(September 2022), 101449. <https://doi.org/10.1016/j.csl.2022.101449>.
- [57] Oesterreich, T. D., Anton, E., Schuir, J., Brehm, A., & Teuteberg, F. (2022). How can I help you? Design principles for task-oriented speech dialog systems in customer service. Information Systems and E-Business Management. <https://doi.org/10.1007/s10257-022-00570-7>.
- [58] Jannach, D. (2022). Evaluating conversational recommender systems: A landscape of research. In Artificial Intelligence Review (Vol. 56, Issue 3). Springer Netherlands. <https://doi.org/10.1007/s10462-022-10229-x>.
- [59] Liu, B., Zhang, P., Shu, Y., Guan, Z., Lu, T., Gu, H., & Gu, N. (2022). Building a Personalized Model for Social Media Textual Content Censorship. Proceedings of the ACM on Human-Computer Interaction, 6(2 CSCW). <https://doi.org/10.1145/3555657>.
- [60] Cheng, L., Jia, W., & Yang, W. (2023). Capture Salient Historical Information: A Fast and Accurate Non-autoregressive Model for Multiturn Spoken Language Understanding. In ACM Transactions on Information Systems (Vol. 41, Issue 2). <https://doi.org/10.1145/3545800>.
- [61] Bhaskaran, S. K., Sreejith, C., & Rafeeqe, P. C. (2018). Neural networks and conditional random fields-based approach for effective question processing. Procedia Computer Science, 143, 211–218. <https://doi.org/10.1016/j.procs.2018.10.381>.
- [62] Tang, Y., Han, H., Yu, X., Zhao, J., Liu, G., & Wei, L. (2021). An Intelligent Question Answering System based on Power Knowledge Graph. IEEE Power and Energy Society General Meeting, 2021-July, 1–5. <https://doi.org/10.1109/PESGM46819.2021.9638018>.
- [63] Wang, C., Dai, S., Wang, Y., Yang, F., Qiu, M., Chen, K., Zhou, W., & Huang, J. (2022). AROBERT: An ASR Robust Pre-Trained Language Model for Spoken Language Understanding. IEEE/ACM Transactions on Audio Speech and Language Processing, 30, 1207–1218. <https://doi.org/10.1109/TASLP.2022.3153268>.
- [64] Fernández-Martínez, F., Luna-Jiménez, C., Kleinlein, R., Griol, D., Callejas, Z., & Montero, J. M. (2022). Fine-Tuning BERT Models for Intent Recognition Using a Frequency Cut-Off Strategy for Domain-Specific Vocabulary Extension. Applied Sciences (Switzerland), 12(3). <https://doi.org/10.3390/app12031610>.
- [65] Quamar, A., Özcan, F., Miller, D., Moore, R. J., Niehus, R., & Kreulen, J. (2020). Conversational BI: An Ontology-Driven Conversation System for Business Intelligence Applications. Proceedings of the VLDB Endowment, 13(12), 3369–3381. <https://doi.org/10.14778/3415478.3415557>.
- [66] Iovine, A., Narducci, F., Musto, C., de Gemmis, M., & Semeraro, G. (2023). Virtual Customer Assistants in finance: From state of the art and practices to design guidelines. Computer Science Review, 47, 100534. <https://doi.org/10.1016/j.cosrev.2023.100534>.
- [67] Razumovskaia, E., Glavaš, G., Majewska, O., Ponti, E. M., Korhonen, A., & Vulic, I. (2022). Crossing the Conversational Chasm: A Primer on Natural Language Processing for Multilingual Task-Oriented Dialogue Systems. Journal of Artificial Intelligence Research, 74, 1351–1402. <https://doi.org/10.1613/JAIR.1.13083>.
- [68] Ruz, G. A., Henríquez, P. A., & Mascareño, A. (2020). Sentiment analysis of Twitter data during critical events through Bayesian network classifiers. Future Generation Computer Systems, 106, 92–104. <https://doi.org/10.1016/j.future.2020.01.005>.
- [69] Matějů, L., Griol, D., Callejas, Z., Molina, J. M., & Sanchis, A. (2021). An empirical assessment of deep learning approaches to task-oriented dialog management. Neurocomputing, 439, 327–339. <https://doi.org/10.1016/j.neucom.2020.01.126>.
- [70] Anantrasirichai, N., & Bull, D. (2022). Artificial intelligence in the creative industries: a review. In Artificial Intelligence Review (Vol. 55, Issue 1). Springer Netherlands. <https://doi.org/10.1007/s10462-021-10039-7>.
- [71] Bidirectional LSTM with attention mechanism and convolutional layer for text classification.
- [72] Mohan, A. T., & Gaitonde, D. V. (2018). A Deep Learning-based Approach to Reduced Order Modeling for Turbulent Flow Control using LSTM Neural Networks. <http://arxiv.org/abs/1804.09269>.
- [73] Graves, A. (2012). Long Short-Term Memory. In: Supervised Sequence Labelling with Recurrent Neural Networks. Studies in Computational Intelligence, vol 385. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-24797-2_4.
- [74] Chen, Q., Zhuo, Z., & Wang, W. (2019). Bert for joint intent classification and slot filling. arXiv preprint arXiv:1902.10909.
- [75] Castellucci, G., Bellomaria, V., Favalli, A., & Romagnoli, R. (2019). Multi-lingual Intent Detection and Slot Filling in a Joint BERT-based Model. Id. <http://arxiv.org/abs/1907.02884>.
- [76] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 2017-Decem (Nips), 5999–6009. <https://doi.org/https://doi.org/10.48550/arXiv.1706.03762>.
- [77] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. NAACL HLT 2019 – 2019. Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference, 1(Mlm), 4171–4186.
- [78] Fernández-Martínez, F., Luna-Jiménez, C., Kleinlein, R., Griol, D., Callejas, Z., & Montero, J. M. (2022). Fine-Tuning BERT Models for Intent Recognition Using a Frequency CutOff Strategy for Domain-Specific Vocabulary Extension. Applied Sciences (Switzerland), 12(3). <https://doi.org/10.3390/app12031610>.
- [79] Sokolova, M., Japkowicz, N., Szpakowicz, S. (2006). Beyond accuracy, F-score and Roc: A family of discriminant measures for performance evaluation. Lecture Notes in Computer Science, 1015–1021. https://doi.org/10.1007/11941439_114.
- [80] Purwandari, K., Cenggoro, T. W., Chanlyn Sigalingging, J. W., & Pardamean, B. (2023). Twitter-based classification for integrated source data of weather observations. IAES International Journal of Artificial Intelligence (IJ-AI), 12(1), 271. <https://doi.org/10.11591/ijai.v12.i1.pp271-283>.
- [81] Palai, P., Agrawal, K., Prasad Mishra, D., & Reddy Salkuti, S. (2023). Text grouping: A comprehensive guide. IAES International Journal of Artificial Intelligence (IJ-AI), 12(3), 1476. <https://doi.org/10.11591/ijai.v12.i3.pp1476-1483>.