

# Survey on Answer Validation for Indonesian Question Answering System (IQAS)

**Abdiansah Abdiansah**

Universitas Sriwijaya, South-Sumatera, 30662, Indonesia  
E-mail: [abdiansah@unsri.ac.id](mailto:abdiansah@unsri.ac.id)

**Azhari Azhari and Anny K. Sari**

Universitas Gadjah Mada, Yogyakarta, 55281, Indonesia  
E-mail: {arison, a\_kartikasari}@ugm.ac.id

Received: 15 April 2017; Accepted: 15 September 2017; Published: 08 April 2018

**Abstract**—Research on Question Answering System (QAS) has been done mainly in English. Unfortunately, for Indonesian, it is still rarely explored whereas Indonesian is the official language used more than 250 million people. Research in the area of Indonesian Question Answering System (IQAS) began in 2005s, and since then only few number of IQAS have been developed. One of the important issues in IQAS is Answer Validation (AV), which is a system that can determine the correctness of QAS. To identify the future scope of research in this area, the need of comprehensive survey on IQAS and AV arises naturally. The goals of this survey are to find the cutting-edge method used in AV and to prove that AV has not been implemented on IQAS. Based on the results, we suggest new opportunities and research challenges for IQAS community.

**Index Terms**—Question Answering System, Indonesian Question Answering System, Answer Validation, QAS, IQAS, AV.

## I. INTRODUCTION

Information Retrieval (IR) is a technique used to seek information from a collection of documents where the result is a list of information that is relevant to the user's keywords [1]. This technique is commonly used by search engines such as Google, Bing, Yahoo, and others. Although the results of the search engine are quite good, sometimes users want to obtain answers quickly and directly without open the links that provided by the search engine [2][3]. Fortunately, research in Natural Language Processing (NLP) developed a system known as Question Answering System (QAS) which is a system processing queries in natural language and can provide direct answers to the user [4].

The architecture of QAS is generally based on pipeline architecture where the data are processed serially so that the output of one stage will be input for the next stage. There are three steps used in QAS [5], namely: 1) question analysis; 2) document retrieval; and 3) answer extraction. Nonetheless, the performance of the

architecture is limited by the dependencies between modules and error propagation. The Community of Natural Language Processing (CNLP) proposed a solution to solved these problems by using Answer Validation (AV) which is a system that can determine the correctness of QAS. This automatic AV is expected to be useful to improve QAS performance, to help humans in the assessment of QAS output, to improve systems confidence self-score, and to develop better criteria for collaborative systems. Systems must emulate human assessment of QAS responses and decide whether an answer is correct or incorrect according to a given snippet. Furthermore, through the first conference of Answer Validation Exercise (AVE) in 2006, AV has been used as a new component for QAS architecture [6].

AVE Conferences have become the main reference for conducting Answer Validation studies. The meetings were held three times in 2006-2008 [6–8] which proposed Recognition Textual Entailment (RTE) [9] as the primary approach where the hypotheses have been built semi-automatically turning the questions and the answers into an affirmative form. Participant systems must return a value YES or NO for each pair of text-hypothesis to indicate if the text entails the hypothesis or not. System results are evaluated against QAS human assessments. Participant systems receive a set of text-hypothesis pairs that were built from QAS main track responses of the CLEF 2006. The methodology for creating these collections was described in [6].

Although RTE is used as the primary approach in the AVE's conference, it still requires analysis of other methods. The purposes of this analysis are to determine whether RTE is the best method for answers validation and to know the other methods that have been used before. There are some researchers using method other than RTE such as Rule-Based [10], Similarity Computation [11], Pipeline Processing [12], and Machine Learning approach [13]. Mapping previous studies is required to know the latest and the best techniques in AV.

The rest of the paper is structured as follows: the next section describes related work of the research. Section 3 describes a brief of QAS, generic QAS architecture, and

system evaluation. Section 4 describes the methodology that is used in the survey. Section 5 and 6 show some related works with a brief overview of the state-of-the-art both AV and IQAS respectively. Section 7 gives some discussions, and the last section draws some conclusions and presents future work proposals.

## II. RELATED WORK

Research in the field of QAS was started in 1965 by [14] researcher who conducted a survey of Frequently Asked Question (FAQ) in English. The survey discussed the basic principles of QAS and the operation methods with a broad scope, including the user interface, database structure, and the techniques used to seek answers from text. At first, QAS is made to respond user questions using existing knowledge in databases as a primary source of information. Although it looks simple, the approach can provide a conceptual model for QAS's applications. The model provides an overview regarding the structure of database and the techniques to find answers or information of a question.

Currently, beside research in English, QAS is also still actively developed in other languages, e.g., Chinese, Arabic, German, etc. [3][4]. Unfortunately for Indonesian, it is still rarely explored whereas Indonesian is the official language used more than 250 million people. There are some scientists that have studied Indonesian Question Answering System (IQAS) [15–29], but the results of the studies remain many issues. One of the crucial problems is validating the answer of IQAS. Therefore, this paper examines the methods that have been used in AV and the potential to be applied on IQAS.

Researches focus on answer validation of QAS started in 2006 in AVE 2006 conference [6]. The objective of these activities is to develop a system that can decide or check QAS's responses whether they are true or false. The goal of AVE is similar to the problem in RTE. Therefore, the scientists are challenged to detect whether or not the text from QAS can be entailed by text from documents. Dataset used in AVE contain three tuples, namely: question, text, and answer hypothesis. The system should provide responses, YES and NO. Researchers in AVE became active in 2006-2007, but this topic is still explored [30–32].

Few studies of IQAS cause difficulties to retrieve information from journals or proceedings. The searching results are taken only from trustworthy sources and can be validated with other sources such as indexing machines (e.g. Google Scholar profiler) and publisher's website. Research on IQAS became active in 2005s conducted by [15] in Cross Language Evaluation Forum (CLEF) on the track of Question Answering (QA). Recent publications obtained while doing literature review was conducted by [29] which focused on IQAS for medical domain using question generation.

## III. QUESTION ANSWERING SYSTEM (QAS)

In the late 1990s, the study of QAS became more active because it was proposed as the main topic in QA Track [33]. The track is an activity used to evaluate the progress of research and development of QAS. Along with many scientists developed QAS in 1995s so that both CLEF and NTCIR also involved as evaluation tools. At that time, the focus and challenges were to find a text containing answers given in a large document. The results of the assessment showed that these challenges could be solved using Information Retrieval, Artificial Intelligence (AI), Natural Language Processing (NLP) techniques and their combination.

Currently, QAS is evaluated directly by TREC. Some evaluations focus on factoid question, i.e. a question that requires a simple fact, entity, or property that are expressed in a few words, e.g., "who was the first president of the united states?". Furthermore, TREC also conducts some evaluations on question types such as definition, e.g., "why does it rain?", list, e.g., "are all of the song written by Michael Jackson?", and relationship, which tries to find a relationship between two entities, e.g., "what is the relationship between bird and chicken?". The evaluation given by NTCIR and CLEF is similar to TREC but it focuses on development of QAS for Asian and European languages. They propose to use other than English language by introducing the term monolingual and cross-lingual. The monolingual aims to develop a QAS using questions and documents in the same language, in contrast to the cross-lingual which uses questions and documents in different languages, e.g., the user submits a question in Spanish, the system will find the answers in English documents. Finally, the answer will be given in Spanish.

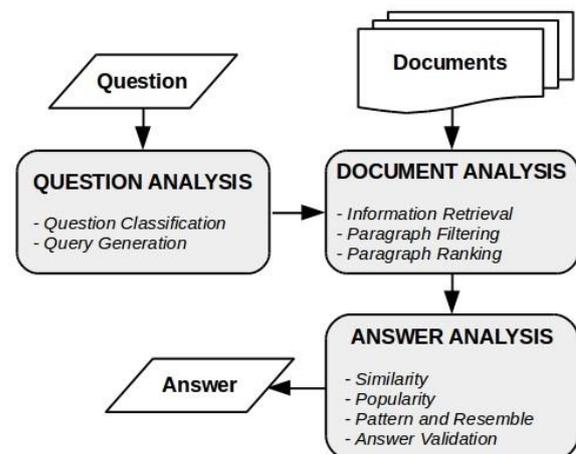


Fig.1. Generic QAS Architecture

### A. Generic QAS Architecture

The process of QAS has three components, namely: 1) question analysis: parsing, question classification, and query reformulation; 2) document analysis: extract candidate documents, identify answers; and 3) answer analysis: extract candidate answers and rank the best one

[2]. These all components work together processing questions and documents at different levels until the final answer is obtained. For example, if the results of question analysis are bad, the result of answers analysis will be certainly bad. Similarly, if the result of question analysis is good, the result of answer analysis will be not necessarily a good one. Therefore, many scientists only focus on one component of QAS. Figure 1 illustrates generic QAS architecture and the below is a brief explanation for each component [34].

### B. QAS Evaluation

The most important test collections currently available have been generated from the data and results of QAS evaluations developed at the TREC conferences organized by the U.S. National Institute of Standards and Technology (NIST), the TREC conferences focuses on various tasks related with IR. For each task, an evaluation method and a particular corpus are defined such that all the participants should provide the result of their systems on the general corpus. NIST evaluates the submissions of the members and organized the conferences where the results are discussed.

TREC evaluation metrics and methodology have become a standard in the field and have been adopted as a reference in other evaluations in QA such as CLEF, NTCIR, or TAC. The first assessment measure employed is the mean reciprocal rank (MRR) [35]. When several ranked answers are allowed, each question is scored according to the inverse of the position of the first passage that contains the correct answer. If none of the passages contain the answer, the question score is 0. The MRR is computed as the mean of the scores of all questions. This measure becomes a standard test for the evaluation of QAS. Later, when only one answer can be returned per question, the MRR could not be used, and the chosen measure is a variation of the answer accuracy called the confidence-weighted score. Given a list of  $Q$  questions ranked according to the confidence of the system to find the correct answer, the confidence-weighted score is:

$$\frac{1}{Q} \sum_{i=1}^Q \frac{\text{number correct in first } i \text{ ranks}}{i} \quad (1)$$

Apart from factoid questions, the inclusion of new types such as list and other questions provokes separate evaluations with different metrics for each of the question types. Systems are given a final score by combining the results obtained for each of the different types. Factoid questions are evaluated by their accuracy i.e. the percentage of questions that has a correct answer. List of questions is assessed using a well-known measure within the area of IR: the F-score combines the recall and precision. Given  $S$  target answers,  $N$  answers are returned by the system, and  $D$  answers are returned that belongs to the target answers, recall is  $R = D/S$  and precision is  $P = D/N$ . The measure of the question is:

$$F = \frac{2 \times R \times P}{R + P} \quad (2)$$

To evaluate “other” questions, the judges are asked to determine a set of minimal pieces of information that should appear in the definition, the so-called information nuggets. The Nuggets are classified as “vital” if they must appear in the answer, and “non-vital” if their appearance in the answer is acceptable. For more information see [36].

## IV. SURVEY METHODOLOGY

The survey was conducted in two topic areas, i.e., AV and IQAS. Therefore, different approaches are used to evaluate both topics. The stages are used for answer-validation’s survey as follows: 1) Looking for research publications related to the answer validation resulted from Elsevier, IEEE, and Google Scholar; 2) Reviewing these researches by analyzing: languages, dataset, methods, and measuring tools; 3) Classifying these researches into research group based on methods used, and 4) Giving a summary associated with the best methods that can be used to validate the answer of QAS. The result of this survey is a map (in tabular data) of answer validation techniques that was used previously and the best method that could be achieved.

Furthermore, the stages are used for IQAS’s survey as follows: 1) Looking for research publications related to the Indonesian Question Answering System (IQAS) resulted from Elsevier, IEEE, and Google Scholar; 2) Reviewing these researches by analyzing: components and finding-answer methods that have been used to build IQAS; 3) Classifying these researches into research group based on research focus, and 4) Giving a summary associated with research opportunities for IQAS. The result of this survey is a proof that the investigation on answer validation for IQAS had never been doing before, therefore it needs to be explored further.

## V. STATE OF THE ART IN ANSWER VALIDATION

Based on the research publication ranging from 2003 to 2015, there are 21 articles that are relevant to the topic of answer validation of QAS. Table 1 shows the list of publication related to the topic sorted by the year of publication. The table has six columns as follows: 1) Refs., research references; 2) Languages, scientists can be used more than one languages for their experiments, for example in [37] and [12]; 3) Datasets; 4) Methods, the methods are used to validate the answers; 5) Groups, contains approaches based on methods that are classified; and 6) Measuring tools are used to measure the system performances. There are three kinds of measuring tools, namely: precision, accuracy, and F-measure.

There are five approaches used to classify the methods, namely: 1) Rule-Based (RB); 2) Similarity Computation (SC); 3) Pipeline Processing (PP); 4) Machine Learning (ML); and 5) Recognition Textual Entailment (RTE).

Grouping the AV's methods into an approach is based on the idea that methods which have similar computational processes may be assigned to the same group. For example, in [38] they used Logic Form Reasoning (LFR), and in [37] they used Logic provers (LP), both used rule-based although the techniques used were different.

Therefore, they can be grouped into Rule-Based (RB) approach. Figure 2 illustrates the methods that are classified into the approaches. Next paragraph explains briefly review associated with researches in Table 1 and groups them into one paragraph per approach.

Table 1. Group of Answer Validation publication based on their approaches

Refs.	Year	Languages	Datasets	Methods	Measuring Tools	Groups
[38]	2003	China	dataset	Logic Form Reasoning	-	RB
[39]	2005	China	dataset	Similarity & Corelation Calculation	74.3% Precision	SC
[40]	2005	English	AVE	Distance-based Approach	37.76% F-Measure	SC
[6]	2007	AVE 2006 Report*			32% Precision	RTE
[43]	2007	Frank	AVE 06	EAT, NER, Decision Algorithm	55% F-Measure	PP
[7]	2007	AVE 2007 Report*			44% Precision	RTE
[44]	2007	Spain	AVE 06	SVM	60% F-measure	ML
[37]	2007	English/Spain	dataset	COGEX (Logic Prover)	English: 43.93% F-measure, Spain: 60.63% F-Measure	RB
[45]	2008	English	RTE-3	RTE	68.75% Accuracy	RTE
[8]	2009	AVE 2008 Report*			54% Precision	RTE
[46]	2009	Spain	SPARTE	RTE	52% Precision	RTE
[42]	2009	English	TREC	Similarity Computation	65% Accuracy	SC
[48]	2010	English	AVE 08	RTE (WordNet, NER)	67% Precision	RTE
[12]	2010	English/Spain	ResPubliQA	EAT, NER, Acronym Checking	English: 65% Accuracy Spain: 57% Accuracy	PP
[47]	2010	Spain	CLEF 06	RTE	53% Accuracy	RTE
[10]	2011	Germany	CLEF 11	Rule-set	44% Accuracy	RB
[13]	2011	Frank	Web	Decision Tree Combination	48% MRR	ML
[41]	2012	Germany	CLEF-QA	LogAnswer Framework	61% correct top rank	SC
[49]	2013	English	AVE 08	RTE	58% Precision 22% F-Score	RTE
[11]	2013	English	CLEF 11	Semantic Similarity	45% Precision	SC
[30]	2013	Russian	ROMIP	RTE	70,4% F-Measure	RTE
[31]	2015	English	Sem-Eval 2015	Distance-based Approach	62,2% Accuracy, 46,07 F-Score	SC
[32]	2016	English	Yahoo! StackOverflow	SVM	62,65% F-Score	ML

\*) These are reports from AVE conferences that contains researches focuses on AV using RTE approach

In Rule-Based (RB) approach, [38] ] presented a logic approach toward answer validation in Chinese Question Answering (CQA). The idea of logic form representation has been used successfully in English QA. Their works extended the Logic Form (LF) representation for Chinese and extracting knowledge from HowNet databases. The answer validation algorithm based on Logic Form Transformation (LFT) was used to validate candidate answer. To testify the validity of the approach to QA answer validation, they borrowed from TREC's testing system since there is still no standard evaluation system for Chinese QA. After experiments, they found that most inaccuracies occurred in LFT were caused by parsing error. In parsing the 631 sentences, the parser generated 586 output trees and failed on remaining 45 sentences. Therefore, a better parser is needed to enhance the precision of logic form transformation. Furthermore, [37]

reported the performance of Language Computer Corporation's natural language logic prover for the English and Spanish subtasks of the Answer Validation Exercise. COGEX was used to take as input in a pair of plain English text snippets, transform them into semantically rich logic forms, automatically generate natural language axioms, and determine the degree of entailment. The system labeled an input pair as positive entailment if its proof score was above the threshold. The approach achieved 43.93% F-measure for the English data and 60.63% on Spanish. Lastly, in RB approach was done by [10] which was building a system that represented questions, answers and texts as formulae in propositional logic derived from dependency structure. They focused on the objective of not using any external resources. The main challenge in [10] was to specify the translation from dependency structure into a logical

representation. Questions were answered by attempting to infer answers from the test documents complemented by background knowledge which was extracted from the background corpora using several knowledge extraction strategies. Result of experiments ran exceed a random baseline, but showed different coverage/accuracy profiles which its accuracy was up to 44% and coverage was up to 65%.

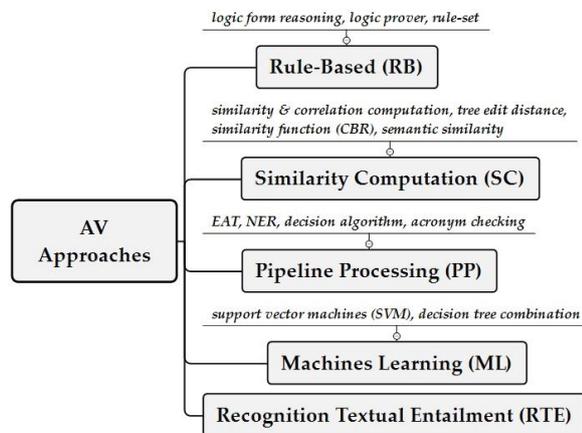


Fig.2. Approaches based on AV's methods

In Similarity Computation (SC) approach, [39] described a Chinese question answering system, which could find answer on the web. The system only utilized the snippet" in the result returned by web search engine as data resource for answer extraction. Answers extraction method was based on the calculation of sentence similarity between question and answer. One novel characteristic of this system was its answer validation method, which combined similarity calculation and correlation calculation to select answer from a number of candidate answers. The experiment result showed that it was more effective than common answer selection methods with total precision of answer was about 74.3%. Next, [40] presented a tree edit distance algorithm applied on the dependency trees of both the text and the hypothesis. The system built is composed by the following modules, namely: 1) a text processing module, for the preprocessing of the input T/H pair; 2) a matching module, which performs the mapping between T and H; and 3) a cost module, which computes the costs of the edit operations. Overall, total accuracy, precision and recall are about 55%, 56% and 50% respectively. In [41][42], they proposed a Case-based Reasoning (CBR) approach to answer validation/answer scoring and reranking in question answering (QA) systems. The approach emphasizes the use of CBR techniques, namely the structural case base, built with annotated MultiNet graphs, and corresponding graph similarity measures. Their experiments showed that the best learned models included CBR features, achieving an MRR up to 0.74 with a correct top-ranked answer shown in 61% of the cases. Lastly, [11] presented a methodology for tackling the problem of answer validation in question answering for reading comprehension tests. The implemented system accepted a document as input and it answered

multiple choice questions about it based on semantic similarity measures. The proposed approach validated the answers, by comparing the text retrieved by Lucene for each question with respect to its candidate answers. The results showed that the proposed system selected the correct answer to a given question with a percentage of 12% more than with a lexical similarity based validation.

In Pipeline Processing (PP) approach, [43] presented a strategy for answer validation in question answering. This strategy was based on their question answering system using step by step pipeline processing. The input of the answer validation was a pair hypothesis-snippet, along with the original question Q and the answer to judge AI. The hypothesis and the text snippet were analyzed by the question answering system, and it defined several criteria which enable to detect whether the snippet justifies the answer. They obtained the following results: precision 0.43, recall 0.84 and F-measure 0.57. Next, [12] applied an IR engine of high performance and a validation step to remove incorrect answers. The IR engine received additional information from the analysis of questions, which produced a slight improvement in results. The mission of answer validation module was to eliminate possible incorrect paragraphs contained in the list returned by the IR engine. Thus, there were more possibilities of giving at the end of the process of a correct answer. The module validated a paragraph when it was considered that the paragraph was correct. If a paragraph was considered as incorrect, the paragraph would be rejected. The experiment result showed accuracy about 50%.

In Machines Learning (ML) approach, [44] presented an entailment relation between entities and tested used they Answer Validation system. The relation as an additional attribute in a SVM setting improved the results of the system close to the best results in the AVE 2006. They used the FreeLing Name Entity Recognizer (NER) to tag numeric expressions (NUMEX), named entities (NE) and time expressions (TIMEX) of both text and hypothesis. Compared with the best in AVE 2006, the system got higher recall and lower precision suggesting that still have room to work on more restrictive filters to detect pairs without entailment. Overall, total precision and recall were about 49% and 50% respectively. Next, [13] developed a QAS based on an answer validation process to be able to handle Web specificity. A large number of candidate answers were extracted from short passages in order to be validated according to question and passages characteristics. The validation module was based on a machine learning approach. It took into account criteria characterizing both passage and answer relevance at surface, lexical, syntactic and semantic levels to deal with different types of texts. They compared the results obtained for factual questions posed on a Web and on a newspaper collection. The system showed outperforms a baseline by up to 48% in MRR using equation (3).

In Recognition Textual Entailment (RTE), [6] reported result of the first Answer Validation Exercise (AVE) at the Cross-Language Evaluation Forum 2006. This task

was aimed at developing systems able to decide whether the answer of a Question Answering system was correct or not. The exercise was described together with the evaluation methodology and the systems results. The starting point for the AVE 2006 was the reformulation of the Answer Validation as a Recognizing Textual Entailment problem, under the assumption that the hypothesis could be automatically generated instantiating hypothesis patterns with the QAS's answers. 11 groups have participated with 38 runs in 7 different languages. The best result was obtained by COGEX in English domain with precision and recall were about 32% and 75%. Next, in AVE 2007, [7] presented the exercise description, the changes in the evaluation methodology with respect to the first edition, and the results of this second edition. The changes in the evaluation methodology had two objectives: the first one was to quantify the gain in performance when more sophisticated validation modules were introduced in QA systems. The second objective was to bring systems based on Textual Entailment to the Automatic Hypothesis Generation problem which was not part itself of the Recognizing Textual Entailment (RTE) task but a need of the Answer Validation setting. Nine groups have participated with 16 runs in 4 different languages. Compared with the QA systems, the results showed an evidence of the potential gain that more sophisticated AV modules introduce in the task of QA. The best result was obtained by DFKI in English domain with precision and recall were about 44% and 71%. Last report from AVE was done by [8] in AVE 2008. They presented the changes in the evaluation with respect to the last edition, and the results of this third edition. The edition wanted to reward AV systems able to detect if all the candidates' answers to a question were incorrect. 9 groups have participated with 24 runs in 5 different languages, and compared with the QA systems, the best results was obtained by DFKI in English domain with precision and recall were about 54% and 78%.

$$MRR = \frac{1}{|Q|} \sum_{i=1}^{|Q|} \frac{1}{rank_i} \quad (3)$$

There are other researches outside of AVE conferences that investigate AV using RTE approach. In [45], they proposed the TE system as an Answer Validation (AV) engine to improve the performance of QAS and help humans in the assessment of QAS's outputs. To achieve these aims and in order to assess the overall performance of TE system and its application in QA tasks, two evaluation environments were presented: Pure Entailment and QA-response evaluation. The former used the corpus and methodology of the PASCAL Recognizing Textual Entailment Challenges. The evaluation environments and the experiments developed were discussed throughout the paper. Next, [46] proposed an external QA ensemble based on answer validation. Like other external ensembles, it did not rely on internal system's features. Nevertheless, it was different from these ensembles in

that: 1) it did not depend on the answer's redundancies, 2) it was not obligated to always select one candidate answer, and 3) it did not only allow returning correct answers but also supported ones. In [47], they proposed a method that allows taking advantage of the outputs of several QAS. This method was based on an answer validation approach that decided about the correctness of answers based on their entailment with a support text, and therefore, that reduced the influence of the answer redundancies and the system confidences. Evaluated over a set of 190 questions from CLEF 2006 collection, the method responded correctly 63% of the questions, outperforming the best QAS participating system (53%) by a relative increase of 19%. Lastly, [48] presented an AV based on Textual Entailment and Question Answering. The important features used to develop the AV system were Named Entity Recognition, Textual Entailment, Question-Answer Type Analysis and Chunk Boundary and Dependency relations. Evaluation scores obtained on the AVE 2008 test set showed precision about 67%. In [49] they presented a rule-based answer validation (AV) system based on textual entailment (TE) recognition mechanism that used semantic features expressed in the Universal Networking Language (UNL). TE system compared the UNL relations in both T and H in order to identify the entailment relation as either validated or rejected. For training and evaluation, the AVE 2008 was used as development set. Obtained 58% precision and 22% F-score for the decision "validated".

## VI. STATE OF THE ART IN INDONESIAN QUESTION ANSWERING SYSTEM

Table 2 shows the list of publication related to the topic of IQAS sorted by year of publication. The table has three columns as follows: 1) Refs. researches reference; 2) IQAS components, contains components are used to build IQAS; and 3) Answer-analysis, contains methods are used to find the answers. Next paragraph explain briefly review associated with researches in Table 2 and grouping them into one paragraph per approach.

In [15–17] they joined an event as participants in the CLEF-2006. They tried to search an answer from set of documents collection. There were 200 documents test from CLEF. In [15] they manually translated the documents from English to Indonesian using Transtool application and separated into paragraphs where one paragraph contains two sentences. They used Monty Tagger to tagging the sentences and Similar Scoring Technique to find final answer. Furthermore, four kinds of labels were used to evaluate the results, namely: W (Wrong), U (Unsupported), X (inexact) and R (Right), in which the results were W=162, U=0, X=36 and R=2. Next, in [16] they translated the documents using ToggleText, a machine translation from web. Gate was used to tag sentences and Lemur as local search engine. Lemur was also used to rank the documents and find final answer. The evaluation results from [16] were W=159, U=13, X=4 and R=14. The results of [16] were better than [15], this can be seen from R's value obtained.

Lastly, research from this group was done by [17] which used generic QAS architecture. The components they used to build IQAS consisted of Question categorization, Document-analysis (passages identification/building [using Lemur and Gate], and passages scoring), and Answer-identification. The evaluation results from [17]

were  $W=175$ ,  $U=1$ ,  $X=4$  and  $R=20$  where the R's value from [17] was increased compared to [16]. Unfortunately, the results of three studies earlier provided low accuracy (below 25.0%). It indicates that it needed more improvement to achieve better results.

Table 2. List of IQAS publication based on their components and answer-analysis's methods

Refs.	Year	Components of IQAS	Methods for Answer-analysis
[15]	2005	Transtools, Tagger (Monty Tagger)	Similar Scoring Technique
[16]	2006	ToogleText, IR (Lemur), Parsing, Tagger (Gate)	Scoring Algorithm
[17]	2007	ToogleText, Question Categorization, Passages Identification/Building (Lemur, Gate), Passages Scoring, Answers Identification	Internet, Word Frequencies and Weighting
[18]	2008	Syntactic Parser, Semantic Analyzer, Question Answering Module	Prolog Rules
[19]	2008	Question Classifier (SVM), Passage Retriever, Answer Finder (SVM)	Machine Learning (SVM)
[20]	2008	Syntactic Parser, Semantic Analyzer, Question Answering Module Augmented with Axioms	Prolog Rules
[21]	2009	OpenEphyra (Framework QAS): Question Analyzer, Query Generator, Search Engine (Lucene), Answer Extractor	Regex, Confidence Score, Support Score
[22]	2011	OpenEphyra (Framework QAS): Question Analyzer, Query Generator, Search Engine, Answer Extractor. Google Translate	Ephyra Answer Finder
[23]	2012	Question Analyzer, Passage Retrieval and Answer Finder	Tf x IDF dan Cosine Similarity
[24]	2012	Question Analyzer, Passage Retrieval and Answer Finder	Factoid (Machine Learning dan NER), Non-Factoid (Pattern Matching dan Semantic Analysis)
[25]	2013	Question Analyzer, Case Retriever and Case Retainer	FreeCBR (Weighted-Euclidean Distance Algorithm)
[26]	2014	Question Analysis, Document Retrieval, Answer Extraction	Rule-Based Method
[27]	2015	Syntax Parsing, Semantic Analysis, Question Generation, QA-Pairs Pattern	Pattern Matching
[28]	2016	question analyzer, passage retrieval, passage scoring, and answer extraction	Scoring Algorithm
[29]	2017	Syntax Parsing, Semantic Analysis, Question Generation, QA-Pairs Pattern	Pattern Matching

In 2007, [18] conducted a study of IQAS focusing on semantic analysis by using three main components, namely: syntactic parser, semantic analyzer and QA module. The semantic representation that was used in the form of logic expression. The expression subsequently made into a rule that would be used by Prolog to finding answers. Furthermore, [20] continued the research of [18] by adding axiom components that can be used as knowledge framework. In general, both studies tried to modeling semantic analysis for Indonesian language.

Studies conducted by [19] and [24] developed IQAS using Machine Learning approach. In [19] they have used components in accordance with generic QAS architecture namely: question classifier, passage retriever and answers finder. The classification process aimed to find Expected Answer Type (EAT) from user's question using three classification methods namely: C4.5, k-Nearest Neighbour (kNN) and Support Vector Machines (SVM). The best result was obtained by SVM with accuracy of 91.97%. Next, Inverse Document Frequency (IDF) was used in passage retriever with f-score value of 0.31% and

the last was YAMCHA, an application based on SVM algorithm was used to find an answer. Overall, the value of MRR (Mean Reciprocal Rank) gained 0.52 to the correct answer. Furthermore, in [24] they built IQAS that has the ability to process queries in the form of factoid and non-factoid question. If the question was a factoid question, it used Named Entity Recognition (NER) and SVM [50][51] to process it. Vice versa, pattern matching and semantic analysis methods were used for non-factoid question. They used the same components as [19] but the difference lied in the answers finding module. The overall result of MRR obtained values were 0.62 for factoid question and 0.80 for non-factoid question.

Furthermore, [21] and [22] developed IQAS using OpenEphyra which was one of the QAS framework that can be modified as needed. Generally, OpenEphyra has same components as the generic QAS architecture, namely: question analysis, question generator, search engine and answer extractor. In [21] they used QAS framework with pattern-based approach to process factoid question. Whereas, in [22] they used QAS

framework for Cross Language Question Answering (CLQA), in which the questions in Indonesian were searched in English document and the answers were given in the Indonesian language.

Beside studied of IQAS with specific approaches as described in the preceding paragraphs, there are scientists who use other approaches such as [23] which used a statistical approach to process non-factoid question. They have used components in accordance with generic QAS architecture, namely: questions analyzer, document retrieval, and answers finder. They also used TfxIDF [52] and cosine similarity techniques to find answers. Next, in [25] they did hybrid system between QAS and Case-based Reasoning (CBR). They used components based on generic QAS architecture but two in three of QAS components were replaced with stages in CBR cycles namely Document-analysis to Case-retrieval, and Answer-analysis to Case-retain, respectively. In [26] they developed closed domain IQAS which searched verse translation of Al-Quran in accordance with user's question. They also used components based generic QAS architecture that consisted of question analysis, document retrieval and answer extraction. Lastly, in [27][29] they built IQAS for medical domain. There were two main components that used for such system i.e. semantic analysis and question generation. Finding the answers was done by generating question-answer pair using pattern matching approach.

## VII. DISCUSSION

Section 2 presenting a brief explanation of QAS, generic architecture and evaluation tools of QAS. It is crucial to know basic of QAS and how to measure the system performance because it will be fundamental for the new scientist on the domain of QAS. Different approaches that have been so far discussed in Section 4 perform fairly well for their domain of scope. Also in Section 5 has discussed methods and components that have been used to build an IQAS. Analysis of two main topics (AV and IQAS) are needed to complement this survey.

### A. Analysis of Answer Validation

Table 1 shows three measurements tools were used in this survey, namely: recall, precision, and accuracy. Each approach gives the different result. Therefore, it cannot be used as comparison. The best approach in Table 1 is determined by the highest value for each measuring tools. The precision is obtained using SC approach about of 74.3% [39]. Furthermore, the accuracy and F-measure are achieved using RTE approach both about of 68.8% [45] and 70.4% [30] respectively. Table 3 shows the highest values for each measuring tool are used.

The precision values listed in Table 3 is quite old [39] which indicates that there is no significant increase in precision measurement. The precision value can be influenced by many things, especially from the dataset and language. Currently, the standard dataset for answer

validation uses TREC (general), NTCIR (Asia) and CLEF (Europe). Table 1 also shows that the highest precision value is obtained using RTE approach [48] about of 67%. The value is held the second position after [39] with the difference is 7.3%. In [39] they do not use the standard dataset to their system, whereas in [48] they were already using the dataset AVE 2008 which is one of dataset standard for answer validation evaluation. Therefore, if the results are viewed from the standard dataset, the result of [48] is more acceptable to be used as the primary reference than the result of [39].

Table 3. The highest values based on measuring tools

Measure Tools	Higher Values	Approaches
Accuracy	68.8%	RTE
Precision	74.3%	SC
F-measure	70.4%	RTE

RTE is the first approach proposed in AVE (2006 – 2008). The evaluation results showed that there was an increase of precision in each conference from 32% (AVE 2006) to 54% (AVE 2008). Furthermore, the value of precision listed in Table 3 has occurred significant increase when compared to the AVE's results which are increased about of 13%. Overall, Table 3 shows that the evaluation results of answer validation are still very low (below 80%) so that there are still rooms to improve these results. The analysis also proved that RTE approach deserves to be explored and suitable for answer validation problem.

Table 4. Summary of IQAS's research groups

No.	Ref.	Research Focus
1	[15-17]	Query Answering – Cross Language Evaluation Forum (CLEF) Cross Lingual QAS
2	[18][20]	Semantic Analysis on IQAS
3	[19][24]	IQAS using Machine Learning
4	[21][22]	Extended and modified of QAS's Frameworks
5	[23][25-29]	Others i.e., Statistical, Case-based Reasoning, Rule-Based, Pattern Matching

### B. Analysis of IQAS

Table 2 shows that researches related to IQAS are still very few. Nevertheless, the research of IQAS relatively stable because it was done almost every year except in 2009. Meanwhile, most researches on IQAS were conducted in 2011. Table 4 shows the result of IQAS's research groups is based on the components used and the topic was undertaken. The table shows that research topics related to answer validation do not yet exist because only small number of researchers are interested in IQAS. Results from previous studies also showed that the methods in Answer-analysis still focus on answer-

finding technique instead of answer validation whereas it is one of the important task in Answer-analysis [4]. Therefore, one of the suggestion to the IQAS community is how to conduct exploration answer validation for Indonesian language.

### VIII. CONCLUSION

In this survey, we present an overview on answer validation for Indonesian Question Answering System (IQAS). We explain about QAS that consist of generic architecture and its evaluation. Also, we describe state of the art of answer validation and IQAS. After that, we conduct some analysis to explore the research challenge related to Answer Validation and IQAS.

Research on QAS has been done mainly in the English language. Unfortunately, for Indonesian, QAS is still rarely explored whereas Indonesian is the official language used more than 250 million people. There are some researchers have been studied on IQAS, but the results of the studies remain many issues. One of the crucial problems is validating the answer of IQAS. Answer Validation (AV) is a system that can determine the correctness of QAS. This automatic AV is expected to be useful to improve QAS performance, to help humans in the assessment of QAS output, to improve systems confidence self-score, and to develop better criteria for collaborative systems. Through the first conference of Answer Validation Exercise in 2006, AV has used as new components for QAS architecture.

The survey conducted on answer validation's research is based on languages, dataset, methods, and measuring tools. Classifying these research into a research group and giving a summary associated with the best methods can be used to validate the answer of QAS. Results of the survey are lists of answer validation techniques used previously. Analysis results AV showed that highest precision values obtained was 74.3% using similarity computation approach, whereas for highest of accuracy and f-measure gained were 68.8% and 70.4% using RTE approach. For the second highest precision values obtained was 67% also using RTE. The survey results demonstrate that RTE approach is feasible to resolve the issue of answer validation.

Few studies of IQAS cause difficulties to retrieve information from journals or proceedings. The searching results are taken only from trustworthy sources, and it can be validated with other sources. The survey was conducted on IQAS's research is based on components that are used to build it. Furthermore, classifying the researches into a research group based on research focus and give a summary associated with research opportunities for IQAS. Results from previous studies also showed that components in Answer-analysis used still focuses on answer finding technique not answer validation. Therefore, one of the research opportunities is exploring the answer validation for IQAS. Previously, it is shown that RTE is a suitable method for answer validation of QAS so that it is probable the method can work well for IQAS.

### REFERENCES

- [1] S. Bütcher, C. L. Clarke, and G. V. Cormack, *Information retrieval: Implementing and evaluating search engines*. MIT Press, 2016. doi:10.1108/02640471111188088.
- [2] S. K. Dwivedi and V. Singh, "Research and Reviews in Question Answering System," in *Procedia Technology*, 2013, vol. 10, pp. 417–424. doi: 10.1016/j.protcy.2013.12.378.
- [3] A. Mishra and S. K. Jain, "A survey on question answering systems with classification," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 28, no. 3, pp. 345–361, 2016. doi:10.1016/j.jksuci.2014.10.007.
- [4] A. Mathur and M. T. U. Haider, "Question answering system: A survey," in 2015 International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials, ICSTM 2015 - Proceedings, 2015, no. May, pp. 47–57. doi:10.1109/ICSTM.2015.7225389.
- [5] Á. Rodrigo and A. Peñas, "On Evaluating the Contribution of Validation for Question Answering," *IEEE Trans. Knowl. Data Eng.*, vol. 27, no. 4, pp. 1157–1161, 2015. doi:10.1109/TKDE.2014.2373363.
- [6] A. Penas, Á. Rodrigo, V. Sama, and F. Verdejo, "Overview of the Answer Validation Exercise 2006," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 5706 LNCS, 2007, pp. 257–264. doi:10.1007/978-3-540-74999-8\_32.
- [7] A. Penas, Á. Rodrigo, and F. Verdejo, "Overview of the Answer Validation Exercise 2007," in *Advances in Multilingual and Multimodal Information Retrieval*, vol. 5706 LNCS, Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 237–248. doi:10.1007/978-3-540-85760-0\_28.
- [8] Á. Rodrigo, A. Peñas, and F. Verdejo, "Overview of the answer validation exercise 2008," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 5706 LNCS, pp. 296–313, 2009. doi:10.1007/978-3-642-04447-2\_35.
- [9] I. Dagan, D. Roth, M. Sammons, and F. Zanzotto, "Recognizing textual entailment: Models and applications," in *Synthesis Lectures on Human Language Technologies*, Morgan & Claypool, 2013, pp. 1–220. doi:10.2200/S00509ED1V01Y201305HLT023.
- [10] S. Babych, A. Henn, J. Pawellek, and S. Pado, "Dependency-based answer validation for German," in *CEUR Workshop Proceedings*, 2011, vol. 1177.
- [11] H. Gómez-Adorno, D. Pinto, and D. Vilariño, "A Question Answering System for Reading Comprehension Tests," in *5th Mexican Conference, MCPR 2013*, 2013, vol. 1087, pp. 354–363. doi:10.1007/978-3-642-38989-4\_36.
- [12] Á. Rodrigo, J. Perez-Iglesias, A. Peñas, G. Garrido, and L. Araujo, "A Question Answering System based on Information Retrieval and Validation," in *Notebook Paper for the CLEF 2010 LABs Workshop*, 2010, pp. 22–23. doi:10.1.1.175.1640.
- [13] A. Grappy, B. Grau, M. H. Falco, A. L. Ligozat, I. Robba, and A. Vilnat, "Selecting answers to questions from Web documents by a robust validation process," in *Proceedings - 2011 IEEE/WIC/ACM International Conference on Web Intelligence, WI 2011*, 2011, vol. 1, pp. 55–62. doi:10.1109/WI-IAT.2011.210.
- [14] R. F. Simmons, "Answering English questions by computer: a survey," *Commun. ACM*, vol. 8, no. 1, pp. 53–70, Jan. 1965. doi:10.1145/363707.363732.
- [15] M. Adriani and Rinawati, "University of indonesia

- participation at query answering-CLEF 2005,” in *CEUR Workshop Proceedings*, 2005, vol. 1171, pp. 4–6.
- [16] S. H. Wijono, I. Budi, L. Fitria, and M. Adriani, “Finding answers to Indonesian questions from English documents,” in *CEUR Workshop Proceedings*, 2006, vol. 1172, pp. 1–4.
- [17] M. Adriani and S. Adiwibowo, “Finding answers using resources in the internet,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 5152 LNCS, 2008, pp. 332–335. doi:10.1007/978-3-540-85760-0-41.
- [18] S. D. Larasati and R. Manurung, “Towards a semantic analysis of bahasa Indonesia for question answering,” in *Proceedings of the 10th Conference of the Pacific Association for Computational Linguistics*, 2007, no. November, pp. 273–280. <http://staf.cs.ui.ac.id/~maruli/pub/pacling07.pdf>.
- [19] A. Purwarianti, M. Tsuchiya, and S. Nakagawa, “A Machine Learning Approach for Indonesian Question Answering System,” in *Proceedings of the 25th Conference on Proceedings of the 25th IASTED International Multi-Conference: Artificial Intelligence and Applications*, 2007, pp. 537–542. <http://dl.acm.org/citation.cfm?id=1295303.1295395>.
- [20] R. Mahendra, S. D. Larasati, R. Manurung, and M. De La Salle Univ, “Extending an Indonesian Semantic Analysis-based Question Answering System with Linguistic and World Knowledge Axioms,” in *Paclic 22: Proceedings of the 22nd Pacific Asia Conference on Language, Information and Computation*, 2008, pp. 262–271. doi:10.1.1.210.120.
- [21] H. Toba and M. Adriani, “Pattern Based Indonesian Question Answering System,” in *Proceedings of the International Conference on Advanced Computer Systems and Information Systems (ICACSIS) University of Indonesia*, 2009.
- [22] M. Iqbal Faruqi and A. Purwarianti, “An Indonesian question analyzer to enhance the performance of Indonesian-English CLQA,” in *Proceedings of the 2011 International Conference on Electrical Engineering and Informatics, ICEEI 2011*, 2011, no. July. doi:10.1109/ICEEI.2011.6021513.
- [23] A. Purwarianti and N. Yusliani, “Sistem Question Answering Bahasa Indonesia untuk Pertanyaan Non-Factoid,” *J. Ilmu Komput. dan Inf.*, vol. 4, no. 1, p. 10, May 2012. doi:10.21609/jiki.v4i1.151.
- [24] A. A. Zulen and A. Purwarianti, “Study and implementation of monolingual approach on Indonesian question answering for factoid and non-factoid question,” in *PACLIC 25 - Proceedings of the 25th Pacific Asia Conference on Language, Information and Computation*, 2011, pp. 622–631. <http://www.scopus.com/inward/record.url?eid=2-s2.0-84863874779&partnerID=40&md5=cfb9e0618d9d103b62f31b3c184750a>.
- [25] A. Fikri and A. Purwarianti, “Case based Indonesian closed domain question answering system with real world questions,” in *7th International Conference on Telecommunication Systems, Services, and Applications, TSSA 2012*, 2012, pp. 181–186. doi:10.1109/TSSA.2012.6366047.
- [26] R. H. Gusmita, Y. Durachman, S. Harun, A. F. Firmansyah, H. T. Sukmana, and A. Suhaimi, “A rule-based question answering system on relevant documents of Indonesian Quran Translation,” in *International Conference on Cyber and IT Service Management (CITSM)*, 2014, pp. 104–107. doi: 10.1109/CITSM.2014.7042185.
- [27] W. Suwarningsih, I. Supriana, and A. Purwarianti, “Discovery Indonesian medical question-answering pairs pattern with question generation,” *Int. J. Appl. Eng. Res.*, vol. 10, no. 14, pp. 34217–34223, 2015.
- [28] M. Z. Naf’an, D. E. Mahmudah, S. J. Putra, and A. F. Firmansyah, “Eliminating unanswered questions from question answering system for Khulafaa Al-Rashidin history,” *Proc. - 6th Int. Conf. Inf. Commun. Technol. Muslim World, ICT4M 2016*, pp. 140–143, 2016. doi:10.1109/ICT4M.2016.33.
- [29] W. Suwarningsih, A. Purwarianti, and I. Supriana, “Reducing the Conflict Factors Strategies in Question Answering System,” in *IOP Conference Series: Materials Science and Engineering*, 2017, vol. 180, pp. 12–20. doi:10.1088/1742-6596/755/1/011001.
- [30] A. Solovyev, “Dependency-Based Algorithms for Answer Validation Task in Russian Question Answering,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 8105 LNAI, 2013, pp. 199–212. doi:10.1007/978-3-642-40722-2\_20.
- [31] I. Nikolova, I. Zamanov, M. Kraeva, N. Hateva, I. Yovcheva, and G. Angelova, “Voltron: A Hybrid System For Answer Validation Based On Lexical And Distance Features,” in *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, 2015, pp. 242–246.
- [32] V. Nguyen and A. Le, “Answer Validation for Question Answering Systems by Using External Resources,” in *Integrated Uncertainty in Knowledge Modelling and Decision Making (Lecture Notes in Computer Science)*, Springer Berlin Heidelberg, 2016, pp. 305–316. doi:10.1007/978-3-319-49046-5\_26.
- [33] P. Ranjan and R. C. Balabantaray, “Question answering system for factoid based question,” in *Contemporary Computing and Informatics (IC3I)*, 2016, pp. 221–224. doi:10.1109/IC3I.2016.7917964.
- [34] N. Indurkha and F. J. Damerau, *Handbook of Natural Language Processing*, 2nd ed. 2010. doi:10.1162/COLI\_r\_00048.
- [35] S. Shekarpour, E. Marx, S. Auer, and A. P. Sheth, “RQUERY: Rewriting Natural Language Queries on Knowledge Graphs to Alleviate the Vocabulary Mismatch Problem,” *AAAI*, pp. 3936–3943, 2017.
- [36] E. M. Voorhees, “Overview of the TREC 2002 question answering track,” in *Proceedings of the 11th Text Retrieval Conference*, 2002, p. 1.
- [37] M. Tatu, B. Iles, and D. Moldovan, “Automatic Answer Validation Using COGEX,” in *7th Workshop of the Cross-Language Evaluation Forum, CLEF 2006*, 2007, vol. 1172, pp. 494–501. doi:10.1007/978-3-540-74999-8\_59.
- [38] Y. Zhang and D. Zhang, “Enabling answer validation by logic form reasoning in Chinese question answering,” in *NLP-KE 2003 - 2003 International Conference on Natural Language Processing and Knowledge Engineering, Proceedings*, 2003, vol. 10, pp. 275–280. doi:10.1109/NLPKE.2003.1275912.
- [39] Dongfeng Cai, Yanju Dong, D. Lv, Guiping Zhang, and Xuelei Miao, “A Web-Based Chinese Question Answering with Answering Validation,” in *2005 International Conference on Natural Language Processing and Knowledge Engineering*, 2005, vol. 1, pp. 499–502. doi:10.1109/NLPKE.2005.1598788.
- [40] M. Kouylekov and B. Magnini, “Recognizing textual entailment with tree edit distance algorithms,” in *PASCAL Challenges on RTE*, 2006, pp. 17–20. <http://>

citeseerx.ist.psu.edu/viewdoc/summary?doi=?doi=10.1.1.124.247.

- [41] I. Glöckner and K. H. Weis, "An integrated machine learning and case-based reasoning approach to answer validation," in *Proceedings - 2012 11th International Conference on Machine Learning and Applications, ICMLA 2012*, 2012, vol. 1, pp. 494–499. doi:10.1109/ICMLA.2012.90.
- [42] S. K. Ray, S. Singh, and B. P. Joshi, "World Wide Web based Question Answering System - a relevance feedback framework for automatic answer validation," in 2009 Second International Conference on the Applications of Digital Information and Web Technologies, 2009, pp. 169–174. 10.1109/ICADIWT.2009.5273942.
- [43] A. L. Ligozat, B. Grau, A. Vilnat, I. Robba, and A. Grappy, "Towards an automatic validation of answers in question answering," in *Proceedings - International Conference on Tools with Artificial Intelligence, ICTAI*, 2007, vol. 2, pp. 444–447. doi:10.1109/ICTAI.2007.156.
- [44] Á. Rodrigo, A. Peñas, J. Herrera, and F. Verdejo, "The Effect of Entity Recognition on Answer Validation," in *7th Workshop of the Cross-Language Evaluation Forum, CLEF 2006, Alicante, Spain, September 20-22, 2006*, 2007, pp. 483–489. doi:10.1007/978-3-540-74999-8\_57.
- [45] O. Ferrandez, R. Munoz, and M. Palomar, "TE4AV: Textual Entailment for Answer Validation," in *2008 International Conference on Natural Language Processing and Knowledge Engineering*, 2008, pp. 1–8. doi:10.1109/NLPKE.2008.4906746.
- [46] A. Tellez-Valero, M. Montes-y-Gómez, L. Villaseñor-Pineda, and A. Peñas, "Improving Question Answering by Combining Multiple Systems Via Answer Validation," in *Computational Linguistics and Intelligent Text Processing*, vol. 4919 LNCS, Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 544–554. doi:10.1007/978-3-540-78135-6\_47.
- [47] A. Tellez-Valero, M. Montes-y-Gómez, and L. Villaseñor-Pineda, "Towards multi-stream question answering using answer validation," *Inform.*, vol. 34, no. 1, pp. 45–54, 2010. doi:10.1.1.61.1041.
- [48] P. Pakray, S. Pal, S. Bandyopadhyay, and A. Gelbukh, "Automatic Answer Validation System on English language," in *2010 3rd International Conference on Advanced Computer Theory and Engineering (ICACTE)*, 2010, vol. 6, pp. V6-329-V6-333. doi:10.1109/ICACTE.2010.5579166.
- [49] P. Pakray, U. Barman, S. Bandyopadhyay, and A. Gelbukh, "Semantic Answer Validation using Universal Networking Language," *Int. J. Comput. Sci. Inf. Technol.*, vol. 3, no. 4, pp. 4927–4932, 2012. doi:10.1.1.259.6069.
- [50] Sharma and K. Sudhir, "Sentiment Predictions using Support Vector Machines for Odd-Even Formula in Delhi," *Int. J. Intell. Syst. Appl.*, vol. 9, no. July, pp. 61–69, 2017. doi:10.5815/ijisa.2017.07.07.
- [51] A. E. Khedr, "Predicting Stock Market Behavior using Data Mining Technique and News Sentiment Analysis," *Int. J. Intell. Syst. Appl.*, vol. 9, no. July, pp. 22–30, 2017. doi:10.5815/ijisa.2017.07.03.
- [52] I. S. I. Abuhaiba and H. M. Dawoud, "Combining Different Approaches to Improve Arabic Text Documents Classification," *Int. J. Intell. Syst. Appl.*, vol. 9, no. April, pp. 39–52, Apr. 2017. doi:10.5815/ijisa.2017.04.05.

## Authors' Profiles



**Abdiansah Abdiansah** received M.Cs (Master of Computer Science) degree in 2008 from Computer Science Department of Universitas Gadjah Mada. Research interests are Question Answering System, Natural Language Processing, Artificial Intelligence, Machine Learning and Expert System.

Email: abdiansah@unsri.ac.id



**Azhari Azhari**, Department of Computer Science and Electronics, Universitas Gadjah Mada. Undergraduate Statistics, Universitas Gadjah Mada, Master of Software Engineering, Institut Teknologi Bandung. Doctor of Computer Science, Universitas Gadjah Mada. Research interests are Intelligent Agent, Software Engineering, Project Management.

Email: arisn@ugm.ac.id



**Anny Kartika Sari**, Department of Computer Science and Electronics, Universitas Gadjah Mada. Undergraduate Mathematics, Universitas Gadjah Mada. M.Sc., University of Twente, Netherlands. Ph.D. La Trobe University, Australia. Research interests are XML-UML mapping, Decision Support System, Information System.

Email: a\_kartikasari@ugm.ac.id

**How to cite this paper:** Abdiansah Abdiansah, Azhari Azhari, Anny K. Sari, "Survey on Answer Validation for Indonesian Question Answering System (IQAS)", *International Journal of Intelligent Systems and Applications (IJISA)*, Vol.10, No.4, pp.68-78, 2018. DOI: 10.5815/ijisa.2018.04.08