

INTISARI

**AUTOMATIC QUESTION GENERATION DARI DOKUMEN TEKS
BAHASA INDONESIA BERDASARKAN LABEL KELAS NON-FACTOID
MENGGUNAKAN NAIVE BAYES CLASSIFIER
(Studi Kasus : Buku Mata Pelajaran Biologi Jenjang SMP)**

Oleh

AMINUDIN

13/357252/PPA/04464

Sistem AQG adalah sistem yang dapat membangkitkan pertanyaan berdasarkan informasi atau teks dokumen dengan menggunakan metode dan pola-pola tertentu. Masalah yang terjadi di dalam pengembangan sistem AQG ini adalah menemukan kecocokan antara pertanyaan yang dihasilkan oleh sistem dengan teks atau dokumen yang di inputkan ke dalam sistem maka dari itu banyak penelitian untuk mengembangkan sistem AQG dengan berbagai macam pendekatan baik pendekatan sintaksis, semantik dan template.

Penelitian ini menggunakan pendekatan *template* dan klasifikasi di dalam membangun sistem AQG dan gabungan *chunking labelling* dan metode *exact matching* untuk menghilangkan kalimat yang tidak sesuai dengan kata target tertentu. Komponen sistem AQG di bagi menjadi beberapa sub komponen yaitu klasifikasi, pengetahuan struktur kalimat dan *template based*. *Naive bayes classification* merupakan bagian dari komponen klasifikasi bertujuan untuk menentukan jenis kalimat definisi, alasan dan metode. Pengetahuan struktur kalimat digunakan untuk mendeteksi pola kalimat menggunakan *chunking labelling* berdasarkan kata target dengan dicocokkan metode *template*.

Berdasarkan hasil pengujian klasifikasi menggunakan *k-fold cross validation* dengan jumlah $k=17$ didapatkan nilai akurasi sebesar 91%, *precision* sebesar 92.22%, *recall* sebesar 93.46%, *f-measures* sebesar 91.92%. Hasil pengujian pembangkitan pertanyaan dengan dataset dokumen Biologi jenjang SMP didapatkan nilai rata-rata masing-masing kelas dengan *precision* sebesar 87%, *recall* sebesar 64% dan *F-measures* 72%.

Kata Kunci : AQG, naive bayes classification, chunking labelling.

ABSTRACT

AUTOMATIC QUESTION GENERATION FROM INDONESIAN TEXT DOCUMENTS BY LABEL CLASS NON-FACTOID USE NAIVE BAYES CLASSIFIER
(Case Study : Subject Book Biology Junior High School)

By

AMINUDIN
13/357252/PPA/04464

AQG system is a system that can generate questions based on the information or text document by using the method and certain patterns. Problems that occur in the development of this AQG system is finding a match between questions generated by the system with text or documents fed into the system and therefore a lot of research to develop a system AQG with a wide range of approaches both approaches syntactic, semantic and template.

This research uses a template approach and classifications in building systems and combined chunking labeling and methods of exact matching to eliminate sentences that do not correspond to a specific target word. AQG system components divided into several sub-components, classification, knowledge of sentence structure and template based. Naive Bayes classification is a component part of the classification aims to determine the type of sentence definition, rationale and methods. Knowledge of sentence structure used to detect phrases using chunking labeling patterns based on the target word with matched template method.

Based on the results of testing the classification using the k-fold cross validation with the number $k = 17$ value obtained an accuracy of 91%, precision 92.22, recall of 93.46, f-measures to 91.92%. The test results with the question of generation dataset Biology SMP document obtained an average of each class with a precision value of 87%, a recall of 64% and 72% F-measures.

Keywords : AQG, naive bayes classification, chunking labelling.