

ABSTRAK

Test of English as a Foreign Language (TOEFL) merupakan salah satu bentuk evaluasi pembelajaran yang membutuhkan kualitas soal yang sangat baik. Pembuatan soal TOEFL dengan cara konvensional tentu menghabiskan banyak waktu dari setiap tahapannya. Teknologi komputer tentu saja dapat digunakan untuk membantu masalah tersebut. Untuk itu, penelitian ini dilakukan untuk membantu menyelesaikan masalah pembuatan soal TOEFL dengan tipe *sentence completion*. Soal-soal ini dihasilkan secara otomatis dari situs berita media asing yang telah dipraproses dan diklasifikasi menggunakan *k-Nearest Neighbor* dengan data soal *sentence completion* dari *ebook* dan buku cetak TOEFL. Selain itu, penentuan pengecoh juga ditentukan berdasarkan atural heuristik untuk meningkatkan kualitas soal. Model sistem ini terdiri dari beberapa tahap, yaitu: (1) Pengumpulan Data (2) Praproses dengan *Natural Language Processing* (3) *POS Tagging* (4) Ekstraksi Fitur Soal (5) Pemisahan dan Pemilihan Kalimat Berita (6) Penentuan dan Pengambilan *Value 7* Fitur (7) Konversi Nilai Data Kategorik (8) Klasifikasi Target Kata Posisi *Blank* (9) Penentuan Heuristik (10) Pemilihan Opsi atau Pengecoh. Hasil evaluasi soal menunjukkan bahwa soal yang dihasilkan memiliki kualitas sangat baik dengan presentase 81.93% setelah dilakukan penilaian oleh *expert*, 81.25% kekonsistenan jawaban, dan 70% kesamaan posisi *blank*.

ABSTRACT

Test of English as a Foreign Language (TOEFL) is one form of a lesson that requires excellent quality questions. TOEFL question's preparation in the conventional way certainly spend a lot of time from each stage. Computer technology can be used to solve the problem. For that thing, this research is done to solve the problem of making TOEFL with sentence completion type. These questions are generated automatically from foreign media news sites that have been processed with Natural Language Processing and classified using k-Nearest Neighbor with sentence completion questions type from the TOEFL ebook and textbook. In addition, the determination of the distractors is also determined based on the heuristic rules to improve the quality of the question. The system model consists of several stages: (1) Data Collection (2) Preprocess with Natural Language Processing (3) POS Tagging (4) Question's Extraction Feature (5) Separation and Selection News' Sentences (6) Determination and Value Taking of 7 Features (7) Conversion of Categorical Data Values (8)) Target Classification for Blank Position Word (9) Heuristic Determination (10) Selection of Option or Distractors. The result of the evaluation shows that the result is very good with the percentage of 81.93% after the expert judgment, 81.25% consistency of the answer, and 70% of the blank position accuracy.

Anita Dyah Pertiwi, 2018

SISTEM PENGHASIL SOAL SENTENCE COMPLETION DALAM TOEFL MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOR DAN NATURAL LANGUAGE PROCESSING

Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu