

INTISARI

oleh:

Hindayati Mustafidah
15/392462/SPA/00556

Tes sebagai instrumen asesmen disajikan dalam bentuk butir-butir soal yang diberikan kepada siswa. Butir soal yang disusun secara tidak langsung ikut menentukan kualitas siswa secara kognitif dalam menguasai materi pelajaran yang diberikan. Oleh karena itu, butir soal perlu untuk dianalisis apakah telah memenuhi standar tingkat kesulitan sebagaimana butir-butir soal Ujian Nasional (UN) sebagai salah satu tolok ukur kualitas soal. Selama ini, untuk mengetahui tingkat kesulitan butir soal, penyusun soal harus melakukan uji coba dengan cara soal diujikan ke siswa, kemudian skor hasil pekerjaan siswa dikumpulkan untuk dianalisis. Hal ini dirasa kurang efisien karena ada kekhawatiran soal akan bocor. Selain itu, meskipun telah terdapat beberapa alat bantu berupa perangkat lunak komputer untuk menganalisis butir soal, namun para guru masih kesulitan dalam mengoperasikan perangkat lunak tersebut dan menginterpretasikan hasil keluarannya. Permasalahan ini dapat diatasi dengan melakukan analisis tingkat kesulitan butir soal menggunakan **model komputasi cerdas**.

Model bekerja dengan dua proses utama yaitu *text processing* dan *case similarity*. *Text processing* melakukan ekstraksi terhadap teks butir soal menggunakan kaidah-kaidah dalam pemrosesan bahasa alami dan digunakan untuk mengidentifikasi setiap teks butir soal untuk diklasifikasikan ke dalam Taksonomi Bloom Revisi (TBR). Proses identifikasi ini menghasilkan kelas kata kerja sebagai dimensi proses kognitif dan kelas kata benda sebagai dimensi pengetahuan yang terkandung dalam teks butir soal. Sementara itu, *case similarity* digunakan untuk mencocokkan setiap hasil klasifikasi dalam TBR dari setiap butir soal terhadap butir-butir soal Ujian Nasional sebagai soal terstandar.

Model berhasil melakukan prediksi tingkat kesulitan butir soal berdasarkan teksnya. Tingkat kesulitan butir soal yang dimaksud adalah meliputi tingkat keterampilan berpikir rendah (*Low Order Thinking Skills/LOTS*) atau tinggi (*High Order Thinking Skills/HOTS*), dan kategori kognitif yang mengacu pada TIMSS (*the Trends in International Mathematics and Science Study*) yaitu *knowing* atau kategori mudah, *applying* atau kategori sedang, dan *reasoning* atau kategori sulit. Akurasi model yang dikembangkan ditentukan berdasarkan perbandingan antara jumlah butir soal UN sebagai soal yang sudah terstandar yang dianalisis oleh model dengan jumlah butir soal UN yang divalidasi oleh tiga orang pakar dan menghasilkan akurasi sebesar 89,42%.

Kata-kata kunci: model, komputasi cerdas, *text processing*, *case similarity*, tingkat kesulitan.

ABSTRACT

by:

Hindayati Mustafidah

15/392462/SPA/00556

The test as an assessment instrument is presented as questions given to students. The items prepared indirectly determine the quality of students cognitively in mastering the given subject matter. Therefore, the items need to be analyzed to determine whether they meet the standard level of difficulty as the items on the National Examination (UN) as one of the benchmarks for the quality of the questions. So far, to determine the level of difficulty of the items, the question compilers have to do a trial run using the questions being tested on students. Scores of student work are collected for analysis. This is considered less efficient because there are concerns about leaks. In addition, even though there have been several tools in the form of computer software to analyze the items, the teachers still have difficulties operating the software and interpreting the output results. This problem can be overcome by analyzing the items' difficulty levels using an intelligent computing model.

The model works with two main processes: text processing and case similarity. Text processing extracts the text of the item items using the principles of natural language processing. It is used to identify each item text to be classified into the Revised Bloom's Taxonomy (TBR). This identification process produces a class of verbs as a dimension of cognitive processes and a class of nouns as a dimension of knowledge contained in the text of the items. Meanwhile, case similarity is used to match each classification result in the TBR of each item to the items on the National Examination as standardized questions.

The model predicted the items' difficulty levels successfully based on the text. The level of difficulty of the items in question includes low (Low Order Thinking Skills/LOTS) or high (High Order Thinking Skills/HOTS) levels and cognitive categories that refer to TIMSS (the Trends in International Mathematics and Science Study), namely knowing or easy category, applying or medium category, and reasoning or difficult category. The model's accuracy was determined based on a comparison between the number of UN items as standardized questions analyzed by the model and the number of UN items validated by three experts. It resulted in an accuracy of 89.42%.

Keywords: model, intelligent computing, text processing, case similarity, difficulty level.