

**EKSTRAKSI FITUR KOMPETENSI SOAL *ERROR IDENTIFICATION*
TOEFL MENGGUNAKAN *NATURAL LANGUAGE PROCESSING***

SKRIPSI

Diajukan untuk Memenuhi sebagian dari
Syarat Memperoleh Gelar Sarjana Komputer
Program Studi Ilmu Komputer



Oleh

Muhammad Melvyn Maunti

1904575

PROGRAM STUDI ILMU KOMPUTER
FAKULTAS PENDIDIKAN MATEMATIKA DAN ILMU PENGETAHUAN
ALAM
UNIVERSITAS PENDIDIKAN INDONESIA
BANDUNG
2024

**EKSTRAKSI FITUR KOMPETENSI SOAL *ERROR IDENTIFICATION*
TOEFL MENGGUNAKAN *NATURAL LANGUAGE PROCESSING***

Oleh

Muhammad Melvyn Maunti

NIM 1904575

Sebuah Skripsi yang Diajukan untuk Memenuhi Salah Satu Syarat Memperoleh
Gelar Sarjana Komputer di Fakultas Pendidikan Matematika dan Ilmu
Pengetahuan Alam

© Muhammad Melvyn Maunti 2024

Universitas Pendidikan Indonesia

Agustus 2024

Hak Cipta Dilindungi Undang Undang

Skripsi ini tidak boleh diperbanyak seluruhnya atau sebagian, dengan dicetak
ulang, difoto kopi, atau cara lainnya tanpa izin dari penulis

MUHAMMAD MELVYN MAUNTI

1904575

**EKSTRAKSI FITUR KOMPETENSI SOAL *ERROR IDENTIFICATION*
TOEFL MENGGUNAKAN *NATURAL LANGUAGE PROCESSING***

DISETUJUI DAN DISAHKAN OLEH PEMBIMBING:

Pembimbing I,



Prof. Dr. Lala Septem Riza, M.T.

NIP. 197809262008121001

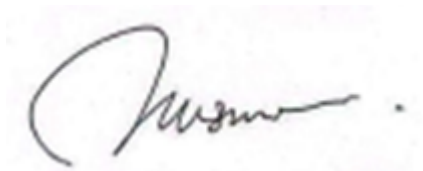
Pembimbing II,



Yaya Wihardi, M.Kom

NIP. 198903252015041001

Mengetahui,
Kepala Program Studi Ilmu Komputer



Dr. Muhammad Nursalman, M.T.

NIP. 197909292006041002

EKSTRAKSI FITUR KOMPETENSI SOAL *ERROR IDENTIFICATION* TOEFL MENGGUNAKAN *NATURAL LANGUAGE PROCESSING*

Oleh

Muhammad Melvyn Maunti — maunti@upi.edu

1904575

ABSTRAK

Tes TOEFL (*Test of English as a Foreign Language*) merupakan salah satu tes *standard* yang digunakan untuk mengukur kemampuan Bahasa Inggris seseorang, terutama bagi mereka yang ingin melanjutkan studi atau bekerja di lingkungan internasional. Salah satu bagian dalam tes TOEFL adalah *Error Identification*, yang menguji kemampuan peserta dalam mengenali kesalahan gramatikal dalam kalimat. Penelitian ini bertujuan untuk membangun sebuah sistem yang mampu mengekstraksi fitur kompetensi gramatikal yang diujikan dalam soal *Error Identification* TOEFL menggunakan teknik *Natural Language Processing* (NLP). Penelitian ini menggunakan 50 soal latihan TOEFL sebagai dataset. Tahap awal penelitian melibatkan pengumpulan data, *preprocessing* teks, dan ekstraksi fitur menggunakan library Spacy dalam bahasa pemrograman Python. Terdapat 13 fitur dan kompetensi gramatikal yang diekstraksi, meliputi *verbs*, *tenses*, *infinitives*, *passives*, *prepositions*, dan lain-lain. Model NLP yang dibangun memanfaatkan informasi POS tagging dan *dependency parsing* untuk mengidentifikasi pola kesalahan gramatikal pada setiap soal. Evaluasi model dilakukan dengan membandingkan hasil ekstraksi fitur otomatis dengan hasil identifikasi manual oleh ahli di bidang Bahasa Inggris. Hasil evaluasi menunjukkan bahwa model mampu mengekstraksi fitur kompetensi gramatikal dengan akurasi rata-rata sebesar 72%. Penelitian ini menunjukkan potensi penerapan NLP dalam mengembangkan sistem otomatis untuk analisis soal TOEFL, khususnya pada bagian *Error Identification*.

Kata Kunci : *TOEFL, Error Identification, Natural Language Processing, Ekstraksi Fitur, Kompetensi Gramatikal.*

TOEFL ERROR IDENTIFICATION QUESTION COMPETENCY FEATURE EXTRACTION USING NATURAL LANGUAGE PROCESSING

Arranged by

Muhammad Melvyn Maunti — maunti@upi.edu

1904575

ABSTRACT

The TOEFL (Test of English as a Foreign Language) test is one of the standard tests used to measure a person's English ability, especially for those who want to continue their studies or work in an international environment. One part of the TOEFL test is Error Identification, which tests participants' ability to recognize grammatical errors in sentences. This study aims to build a system that is able to extract grammatical features and competencies tested in TOEFL Error Identification questions using Natural Language Processing (NLP) techniques. This study uses 50 TOEFL practice questions as a dataset. The initial stages of the study include data collection, text preprocessing, and feature extraction using the Spacy library in the Python programming language. There are 13 grammatical features and competencies extracted, including verbs, tenses, infinitives, passives, prepositions, and others. The NLP model built utilizes POS tagging information and dependency parsing to identify grammatical error patterns in each question. The evaluation model is carried out by comparing the results of automatic feature extraction with the results of manual identification by four experts in the field of English. The evaluation results show that the model is able to extract grammatical features and competencies with an average accuracy of 72%. This study shows the potential of applying NLP in developing an automated system for TOEFL question analysis, especially in the Error Identification section.

Keywords : TOEFL, Error Identification, Natural Language Processing, Feature Extraction, Gramatical Competences

DAFTAR ISI

ABSTRAK	i
ABSTRACT	ii
KATA PENGANTAR.....	iii
UCAPAN TERIMAKASIH.....	iv
DAFTAR ISI.....	vi
DAFTAR TABEL.....	ix
DAFTAR GAMBAR	x
DAFTAR LAMPIRAN	xi
BAB I PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah	5
1.3 Tujuan Penelitian.....	5
1.4 Manfaat Penelitian.....	6
1.5 Batasan Masalah.....	6
1.6 Sistematika Penulisan	6
BAB II KAJIAN PUSTAKA	8
2.1 Tes <i>English as a Second Language</i> (ESL)	8
2.1.1 Konsep Tes ESL.....	8
2.1.2 TOEFL	9
2.1.3 Soal <i>Error Identification</i>	10
2.2 <i>Natural Language Processing</i> (NLP).....	12
2.2.1 Definisi NLP	12
2.2.2 Tahapan Umum dalam NLP.....	13
2.3 Ekstraksi Fitur	22

2.3.1	Definisi Ekstraksi Fitur	22
2.3.2	Peran Penting Ekstraksi Fitur dalam NLP.....	23
2.3.3	Metode Ekstraksi Fitur	25
2.3.4	Ekstraksi Fitur Berbasis Morfologi	26
2.3.5	Ekstraksi Fitur Berbasis Sintaksis	28
BAB III METODE PENELITIAN.....		31
3.1	Desain Penelitian	31
3.2	Metode Penelitian	33
3.2.1	Metode Pengumpulan Data	33
3.2.2	Metode Pengembangan Perangkat Lunak	33
3.3	Alat dan Bahan Penelitian	35
3.3.1	Alat Penelitian	35
3.3.2	Bahan Penelitian.....	36
BAB IV HASIL DAN PEMBAHASAN		37
4.1	Pengumpulan Data.....	37
4.3.1	Data Soal <i>Error Identification</i>	37
4.3.2	Data Fitur dan Kompetensi Soal	39
4.2	Rancangan Model Komputasi	45
4.2.1	Data Collection.....	47
4.2.2	Data Preprocessing	47
4.2.3	Proses Ekstraksi Fitur.....	49
4.2.3	Proses Evaluasi Model	61
4.3	Implementasi Model Komputasi	62
4.3.1	Data Collection.....	63
4.3.2	Data Preprocessing	64
4.3.3	Proses Ekstraksi Fitur.....	66

4.2.3 Proses Evaluasi Model	79
4.4 Rancangan Skenario Eksperimen	79
4.5 Hasil dan Pembahasan Eksperimen	80
4.5.1 Hasil Eksperimen Ekstraksi Fitur.....	80
4.5.1 Analisa Eksperimen.....	86
BAB V KESIMPULAN DAN SARAN.....	95
5.1 Kesimpulan.....	95
5.2 Saran	95
DAFTAR PUSTAKA	97
LAMPIRAN	104

DAFTAR TABEL

Tabel 4.1 Representasi data soal	38
Tabel 4.2 Contoh sampel data soal.....	39
Tabel 4.3 Data fitur dan kompetensi soal <i>error identification</i>	41
Tabel 4.4 Detail variabel input pengguna	47
Tabel 4.5 Proses Tokenisasi dan POS Tagging.....	48
Tabel 4.6 Proses Dependency Parsing	48
Tabel 4.7 Representasi Rubrik Evaluasi.	61
Tabel 4.8 Hasil Ekstraksi Fitur oleh Sistem.....	81
Tabel 4.9 Sampel hasil perhitungan confusion matrix setiap soal	86
Tabel 4.10 Perbandingan Hasil Eksperimen dengan Evaluasi Expert	87
Tabel 4.11 Hasil Evaluasi dari Eksperimen	90
Tabel 4.12 <i>Accuracy score</i> dengan Confusion Matrix dari setiap fitur	91
Tabel 4.13 Perbandingan dengan penelitian sebelumnya	94

DAFTAR GAMBAR

Gambar 2.1 Contoh Proses Tokenisasi	14
Gambar 2.2 Treebank <i>part of speech tagset</i> (Marcus dkk., 1993).....	17
Gambar 2.3 Ilustrasi dari penggunaan <i>Named Entity Recognition</i> (NER).....	19
Gambar 2.4 <i>Dependency parsing</i> untuk kalimat bahasa Inggris.....	20
Gambar 3.1 Diagram Desain Penelitian.....	31
Gambar 3.2 Model Waterfall (Pressman, 2010)	34
Gambar 4.1 Diagram Model Komputasi Sistem Fitur Ekstraksi Soal Error Identification.	46
Gambar 4.2 Diagram ekstraksi fitur.....	50
Gambar 4.3 Visualisasi akurasi model setiap fitur pada rata-rata keempat expert.	92

DAFTAR LAMPIRAN

Lampiran 1 Hasil analisa confusion matrix dari identifikasi manual oleh expert.	104
--	-----

DAFTAR PUSTAKA

- Aggarwal, C. C., & Zhai, C. (2012). A survey of text classification algorithms. *Mining text data*, 163–222.
- Alderson, J. C. (2000a). *Assessing reading*. Cambridge University Press.
- Alderson, J. C. (2000b). *Assessing reading*. Cambridge University Press.
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests* (Vol. 1). Oxford University Press.
- Bengio, Y., Courville, A., & Vincent, P. (2013). Representation learning: A review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence*, 35(8), 1798–1828.
- Biswas, P., Sharan, A., & Kumar, R. (2014). Question Classification using syntactic and rule based approach. *2014 international conference on advances in computing, communications and informatics (ICACCI)*, 1033–1038.
- Blodgett, S. L., Green, L., & O'Connor, B. (2016). Demographic Dialectal Variation in Social Media: A Case Study of African-American English. *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, 1119–1130. <https://doi.org/10.18653/v1/D16-1120>
- Burstein, J., Tetreault, J., & Madanani, N. (2013). *Handbook of Automated Essay Evaluation* (M. D. Shermis & J. Burstein, Ed.). Routledge. <https://doi.org/10.4324/9780203122761>
- Cambria, E., Poria, S., Gelbukh, A., & Thelwall, M. (2017). Sentiment analysis is a big suitcase. *IEEE Intelligent Systems*, 32(6), 74–80.
- Cambria, E., & White, B. (2014). Jumping NLP curves: A review of natural language processing research. *IEEE Computational intelligence magazine*, 9(2), 48–57.

- Chalhoub-Deville, M., & Turner, C. E. (2000). What to look for in ESL admission tests: Cambridge certificate exams, IELTS, and TOEFL. *System*, 28(4), 523–539.
- Chapelle, C. A., & Douglas, D. (2006). *Assessing language through computer technology*. Cambridge University Press.
- Chiang, D. (2007). Hierarchical phrase-based translation. *computational linguistics*, 33(2), 201–228.
- Chiswick, B. R., & Miller, P. W. (2007). *The economics of language: International analyses*. Routledge.
- Chodorow, M., & Leacock, C. (2000). An unsupervised method for detecting grammatical errors. *1st Meeting of the North American Chapter of the Association for Computational Linguistics*.
- Chowdhary, K., & Chowdhary, K. R. (2020). Natural language processing. *Fundamentals of artificial intelligence*, 603–649.
- Cummins, J. (2000). *Language, power, and pedagogy: Bilingual children in the crossfire* (Vol. 23). Multilingual matters.
- Das, S. R., & Chen, M. Y. (2007). Yahoo! for Amazon: Sentiment extraction from small talk on the web. *Management science*, 53(9), 1375–1388.
- De Marneffe, M.-C., & Manning, C. D. (2008). The Stanford typed dependencies representation. *Coling 2008: proceedings of the workshop on cross-framework and cross-domain parser evaluation*, 1–8.
- Dozat, T., & Manning, C. D. (2016). Deep biaffine attention for neural dependency parsing. *arXiv preprint arXiv:1611.01734*.
- Educational Testing Service. (2024). *TOEFL Opening doors to a wide and diverse world*. <https://www.ets.org/toefl.html>
- Ellis, R. (1997). Second language acquisition. *The United States: Oxford*, 98.

- Feldman, R. (2013). Techniques and applications for sentiment analysis. *Communications of the ACM*, 56(4), 82–89.
- Fulcher, G., & Davidson, F. (2007). *Language testing and assessment*. Routledge London and New York.
- Goldberg, Y. (2016). A primer on neural network models for natural language processing. *Journal of Artificial Intelligence Research*, 57, 345–420.
- Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Liu, T., Wang, X., Wang, G., Cai, J., & others. (2018). Recent advances in convolutional neural networks. *Pattern recognition*, 77, 354–377.
- Hadijah, S., Shalawati, S., & Idayani, A. (2018). Analisis kemampuan mahasiswa dalam berkomunikasi secara lisan pada tes simulasi IELTS (International english language testing system). *GERAM: Gerakan Aktif Menulis*, 6(2), 91–102.
- He, Y., Hui, S. C., & Quan, T. T. (2009). Automatic summary assessment for intelligent tutoring systems. *Computers & Education*, 53(3), 890–899. <https://doi.org/https://doi.org/10.1016/j.compedu.2009.05.008>
- Hilke, R., & Wadden, P. (1997). The Toefl and Its Imitators: Analyzing the Toefl and Evaluating Toefl-Prep Texts. *RELC Journal*, 28(1), 28–53. <https://doi.org/10.1177/003368829702800103>
- Hutchinson, B., & Mitchell, M. (2018). *50 Years of Test (Un)fairness: Lessons for Machine Learning*. <https://doi.org/10.1145/3287560.3287600>
- Indurkha, N., & Damerau, F. J. (2010). *Handbook of natural language processing*. Chapman and Hall/CRC.
- Jacobs, H. L., & others. (1981). *Testing ESL composition: A practical approach. English composition program*. ERIC.
- Jurafsky, D., & Martin, J. H. (2008). *Speech and language processing (prentice hall series in artificial intelligence)*. Prentice Hall NJ.

- Kunnan, A. J. (2000). *Fairness and validation in language assessment: Selected papers from the 19th Language Testing Research Colloquium, Orlando, Florida* (Nomor 9). Cambridge University Press.
- Leacock, C., Chodorow, M., Gamon, M., & Tetreault, J. (2014). *Automated Grammatical Error Detection for Language Learners, Second Edition*. Springer International Publishing. <https://doi.org/10.1007/978-3-031-02153-4>
- Lee, J. S. Y. (2009). *Automatic correction of grammatical errors in non-native English text*. Massachusetts Institute of Technology.
- Li, J., Sun, A., Han, J., & Li, C. (2020). A survey on deep learning for named entity recognition. *IEEE transactions on knowledge and data engineering*, 34(1), 50–70.
- Li, X., & Roth, D. (2002). Learning question classifiers. *COLING 2002: The 19th International Conference on Computational Linguistics*.
- Li, X., Roth, D., & Small, K. (2004). The role of semantic information in learning question classifiers. *Proceedings of the International Joint Conference on Natural Language Processing*.
- Liddy, E. D. (2001). *Natural language processing*.
- Liu, B. (2022). *Sentiment analysis and opinion mining*. Springer Nature.
- Manning, C. D. (2008). *Introduction to information retrieval*. Syngress Publishing,.
- Manning, C., & Schutze, H. (1999). *Foundations of statistical natural language processing*. MIT press.
- Marcus, M., Santorini, B., & Marcinkiewicz, M. A. (1993). Building a large annotated corpus of English: The Penn Treebank. *Computational linguistics*, 19(2), 313–330.
- Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams engineering journal*, 5(4), 1093–1113.

- Meystre, S. M., Savova, G. K., Kipper-Schuler, K. C., & Hurdle, J. F. (2008). Extracting information from textual documents in the electronic health record: a review of recent research. *Yearbook of medical informatics*, 17(01), 128–144.
- Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *Advances in neural information processing systems*, 26.
- Mohasseb, A., Bader-El-Den, M., & Cocea, M. (2018). Question categorization and classification using grammar based approach. *Information Processing & Management*, 54(6), 1228–1243.
- Nadeau, D., & Sekine, S. (2007). A survey of named entity recognition and classification. *Linguisticae Investigationes*, 30(1), 3–26.
- Nivre, J. (2005). Dependency grammar and dependency parsing. *MSI report*, 5133(1959), 1–32.
- Pang, B., Lee, L., & others. (2008). Opinion mining and sentiment analysis. *Foundations and Trends® in information retrieval*, 2(1–2), 1–135.
- Porter, M. F. (1980). An algorithm for suffix stripping. *Program*, 14(3), 130–137.
- Pressman, R. S. (2010). A practitioner's approach. *Software Engineering*, 2, 41–42.
- Sahami, M., Dumais, S., Heckerman, D., & Horvitz, E. (1998). A Bayesian approach to filtering junk e-mail. *Learning for Text Categorization: Papers from the 1998 workshop*, 62, 98–105.
- Salton, G., & Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. *Information processing & management*, 24(5), 513–523.
- Sarrouti, M., Lachkar, A., & Ouatic, S. E. A. (2015). Biomedical question types classification using syntactic and rule based approach. *2015 7th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K)*, 1, 265–272.

- Sharpe, P. J. (2004). *Barron's How to Prepare for the TOEFL with CD-ROM, 11th Edition* (Eleventh edition). Barrons Educational Series.
- Shohamy, E. (2020a). *The power of tests: A critical perspective on the uses of language tests*. Routledge.
- Shohamy, E. (2020b). *The power of tests: A critical perspective on the uses of language tests*. Routledge.
- Socher, R., Perelygin, A., Wu, J., Chuang, J., Manning, C. D., Ng, A. Y., & Potts, C. (2013). Recursive deep models for semantic compositionality over a sentiment treebank. *Proceedings of the 2013 conference on empirical methods in natural language processing*, 1631–1642.
- Tetreault, J., Blanchard, D., & Cahill, A. (2013). A report on the first native language identification shared task. *Proceedings of the eighth workshop on innovative use of NLP for building educational applications*, 48–57.
- Vapnik, V. (2013). *The nature of statistical learning theory*. Springer science & business media.
- Weigle, S. C. (2002). *Assessing writing*. Cambridge University Press.
- Yadav, V., Bethard, S., & Surdeanu, M. (2020). Unsupervised Alignment-based Iterative Evidence Retrieval for Multi-hop Question Answering. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 4514–4525. <https://doi.org/10.18653/v1/2020.acl-main.414>
- Yannakoudakis, H., Briscoe, T., & Medlock, B. (2011). A New Dataset and Method for Automatically Grading ESOL Texts. Dalam D. Lin, Y. Matsumoto, & R. Mihalcea (Ed.), *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies* (hlm. 180–189). Association for Computational Linguistics. <https://aclanthology.org/P11-1019>
- Young, T., Hazarika, D., Poria, S., & Cambria, E. (2018). Recent trends in deep learning based natural language processing. *ieee Computational intelligence magazine*, 13(3), 55–75.

Zhang, Y., Jin, R., & Zhou, Z.-H. (2010). Understanding bag-of-words model: a statistical framework. *International journal of machine learning and cybernetics*, 1, 43–52.