

SISTEM IDENTIFIKASI PLAGIAT MENGGUNAKAN KOMBINASI ALGORITMA GENETIKA-SISTEM FUZZY

Mohamad Syahrul Mubarak¹, Adiwijawa², Suyanto³

¹Teknik Informatika, Fakultas Teknik Informatika, Universitas Telkom

Abstrak

Plagiat merupakan tindakan pencurian hasil karya orang lain dan dapat dijumpai salah satunya di dunia perkuliaan. Praktek plagiat yang sering dilakukan mahasiswa adalah copy paste dan paraphrase kalimat. Manusia dengan menggunakan intuisinya dapat melakukan identifikasi plagiat atau tidaknya suatu dokumen. Namun perbedaan intuisi antara manusia yang satu dengan yang lain dapat menghasilkan identifikasi yang berbeda. Selain itu, seiring bertambahnya jumlah dokumen yang diidentifikasi maka semakin bertambah pula waktu yang dibutuhkan. Dalam tugas akhir ini digunakan sistem fuzzy untuk membuat sistem yang dapat mengakuisisi intuisi manusia dalam mengidentifikasi plagiat atau tidaknya suatu dokumen. Semakin optimal sistem fuzzy yang ada maka semakin akurat pula hasil identifikasi yang dilakukan Dengan menggunakan algoritma genetika, sistem fuzzy yang berakurasi tinggi dalam identifikasi plagiat dapat dihasilkan.

Kata Kunci : plagiat, identifikasi, copy paste, paraphrase, sistem fuzzy, algoritma genetika

Abstract

Plagiarism is stealing a work of people. We are able to find this crime in college. The students mostly do copy paste and paraphrase the sources. Using his intuition, human is able to identify plagiarism in the documents. However the intuition of each identifier is different on each other. It can give difference identification result. Furthermore, the increaseing number of documents to be identified increases time of identification process.

This final project uses fuzzy system which is able to create an intuitive system that can work as human intuition to identify the documents whether they are plagiarism or not. An optimum fuzzy system can optimally identify the documents. We use genetic algorithm to create this optimum fuzzy system which gives high accuracy in identification process as human does.

Keywords : plagiarism, identification, copy paste, paraphrase, fuzzy sistem, genetic algorithm

Telkom
University

1. Pendahuluan

1.1 Latar belakang

Plagiat merupakan tindakan pencurian hasil karya orang lain, pelanggaran terhadap hak cipta dan hal ini tidak dibenarkan karena dapat merugikan orang lain. Tindakan plagiat terjadi di semua bidang, salah satunya adalah di dunia pendidikan. Kasus yang paling sering muncul di dunia pendidikan, khususnya perkuliahan, yakni plagiat dalam pengerjaan berbagai tugas, baik tugas harian, tugas besar, bahkan tugas akhir. Terlebih lagi, apabila tugas-tugas tersebut dituangkan dalam bentuk file dokumen.

Akar kata 'plagiat' berasal dari bahasa Latin, '*plagiarius*', yang berarti 'penculik', dan '*plagiare*', berarti 'mencuri' [8]. Kamus Besar Bahasa Indonesia (KBBI) mendefinisikan 'plagiat' sebagai "pengambilan karangan (pendapat) orang lain dan menjadikannya seolah-olah karangan (pendapat) sendiri. Sedangkan menurut wikipedia, plagiarisme atau penjiplakan adalah penggunaan gagasan, informasi, atau tulisan orang lain tanpa memberikan informasi yang cukup tentang sumber aslinya.

Untuk mengidentifikasi adanya plagiat atau tidak pada suatu dokumen melibatkan penggunaan intuisi manusia. Sedangkan intuisi manusia yang satu dengan yang lain dapat berbeda. Dengan kata lain hal ini berkaitan dengan unsur ketidakpastian dan kebenaran parsial. Oleh karena itu dibutuhkan suatu sistem yang dapat melakukan fungsionalitas tersebut dimana sistem ini sanggup memecahkan persoalan yang berhubungan dengan unsur ketidakpastian dan kebenaran parsial.

Sistem fuzzy merupakan salah satu bahasan dalam *soft computing* yang karakteristiknya adalah menangani permasalahan yang berhubungan dengan ketidakpastian dan kebenaran parsial [13]. Sistem fuzzy digunakan untuk membangun sistem identifikasi yang telah disebutkan sebelumnya. Fungsi keanggotaan dan aturan pada sistem fuzzy merupakan bagian-bagian penting yang sangat mempengaruhi tingkat optimasi suatu sistem fuzzy.

Berkaitan dengan optimasi, bahasan lain dalam *soft computing* yang berhubungan dengan hal ini adalah algoritma genetika. Konsep yang penting disini adalah hereditas, yaitu sebuah ide yang menyatakan bahwa sifat-sifat individu dapat dikodekan dengan cara tertentu sehingga sifat-sifat tersebut dapat diturunkan kepada generasi berikutnya [11]. Dalam algoritma genetika dilakukan proses evolusi, seperti teori evolusi pada biologi, yang akan menyeleksi individu-individu yang ada pada suatu generasi sehingga pada akhirnya tersisa/dihasilkan individu terpilih yang dapat bertahan hidup untuk menghasilkan dan mewariskan sifat-sifatnya pada generasi berikutnya. Dapat dikatakan pula bahwa individu terpilih tersebut merupakan individu yang paling tangguh/optimal pada generasinya.

Dengan menggabungkan karakteristik yang dimiliki sistem fuzzy dan algoritma genetika, maka sistem identifikasi plagiat pada dokumen dibangun. Bentuk dan domain fungsi keanggotaan merupakan individu pada algoritma genetika yang akan mengalami proses evolusi, dan akhir dari proses evolusi tersebut dihasilkan bentuk dan domain fungsi keanggotaan optimal. Bentuk dan domain fungsi keanggotaan tidak dapat ditetapkan diawal karena tidak terdapat *a priori information* sehingga untuk keduanya digunakan pengkodean individu secara dinamis [14]. *A priori information* biasanya diperoleh dari seorang ahli yang mengetahui secara persis karakter variabel permasalahan [14]. proses-proses yang dilakukan dalam mengidentifikasi masalah yang akan dicari solusinya.

1.2 Perumusan masalah

Berdasarkan latar belakang tersebut, maka dapat dirumuskan permasalahan sebagai berikut:

1. Menganalisis dan mengimplementasikan penggunaan algoritma genetika dengan pengkodean individu secara dinamis untuk menghasilkan sistem fuzzy yang memiliki akurasi tinggi dalam identifikasi plagiat.
2. Menganalisis dan mengimplementasikan penggunaan system fuzzy yang dihasilkan pada poin pertama diatas dalam mengidentifikasi plagiat pada dokumen.

1.3 Tujuan

Tujuan dari penelitian tugas akhir ini adalah:

1. Menghasilkan dan menganalisis sistem fuzzy menggunakan algoritma genetika dengan pengkodean individu secara dinamis.
2. Membangun dan menganalisis keakuratan sistem identifikasi dokumen yang telah dibangun.
3. Menggunakan sistem yang telah dibangun sebagai alat bantu proses identifikasi plagiat pada dokumen tugas Mata Kuliah Kapita Selekt.

1.4 Batasan masalah

Terdapat beberapa batasan dalam penelitian Tugas Akhir ini, antara lain:

- a. Dokumen yang akan diidentifikasi adalah dokumen yang berekstensi *.txt.
- b. Dokumen yang digunakan adalah dokumen berbahasa Indonesia.
- c. Keseluruhan dokumen berisi topik bahasan yang sama.
- d. Identifikasi plagiat dilakukan antar N dokumen yang ada, dimana N adalah jumlah keseluruhan dokumen.
- e. Jumlah nilai linguistik untuk tiap variabel input fuzzy adalah 3 nilai. Hal ini dimaksudkan untuk memperkecil ruang pencarian solusi sekaligus merepresentasikan nilai linguistik “rendah”, “sedang”, dan “tinggi”.

1.5 Metodologi penyelesaian masalah

Metode yang digunakan untuk menyelesaikan Tugas Akhir ini adalah :

1. Studi Literatur, yakni dengan mempelajari referensi dan literatur yang relevan yang membahas tentang *Soft computing* (sistem fuzzy, algoritma genetika, dan *evolutionary design of fuzzy systems*)
2. Pembangunan model, dengan menggunakan konsep analisis dan desain terstruktur yang divisualisasikan dengan diagram blok.
3. Implementasi, *coding* dengan menggunakan MATLAB R2006a untuk membangun perangkat lunak yang dimaksud berdasarkan model yang telah dibangun pada tahap sebelumnya.
4. Analisa hasil, melakukan pengujian dan analisa dari sistem yang telah dibangun pada tahap implementasi. Pengujian dilakukan dengan dilakukan

dengan mengujikan data uji berupa file-file dokumen (*.txt), kemudian dilakukan analisa dari hasil yang didapatkan tersebut sesuai dengan parameter pengujian.

5. Pembuatan laporan, melakukan pelaporan hasil pengerjaan Tugas Akhir berupa analisis sistem yang dibangun beserta dokumentasinya.



5. Kesimpulan dan Saran

5.1 Kesimpulan

Dari hasil pengujian dan analisa yang telah dilakukan, dapat diambil beberapa kesimpulan sebagai berikut:

1. Fungsi keanggotaan dan aturan fuzzy adalah komponen yang sangat penting pada sistem fuzzy. *Tuning* fungsi keanggotaan dan pendefinisian aturan fuzzy untuk menghasilkan sistem fuzzy yang optimal berdasarkan suatu kasus tertentu biasanya dilakukan oleh pakar yang sesuai dengan kasus yang dihadapi. Namun algoritma genetika sebagai salah satu algoritma optimasi mampu melakukan fungsi *tuning* tersebut sehingga sistem fuzzy yang dihasilkan memiliki akurasi 100% terhadap data testing.
2. Dengan karakteristiknya yang hanya fokus pada solusi yang hampir mendekati solusi absolut selama proses pencarian, algoritma genetika dapat menemukan solusi yang optimal meskipun dengan ruang pencarian yang tidak terlalu besar.
3. Untuk pelatihan Algoritma Genetika dengan menggunakan panjang kromosom 244 bit, jumlah kromosom untuk satu kromosom, dengan peluang pindah silang 0.8 dan peluang mutasi $1/N$ (N = jumlah gen dalam kromosom), yang memberikan hasil yang paling baik adalah dengan menggunakan ukuran populasi (ruang pencarian) 50.
4. Nilai threshold paraphrase 0.5 adalah nilai threshold yang paling baik untuk menghitung prosentase *word alignment* pada dokumen.
5. Bentuk fungsi keanggotaan terbaik untuk variabel input *Sentence Alignment* adalah trapesium dengan nilai linguistik **Rendah** terletak pada interval [0.. 44.93]. Sementara nilai linguistik **Sedang** terletak pada interval [4.412.. 76.53]. Dan nilai linguistik **Tinggi** terletak pada interval [46.39.. 100].
6. Bentuk fungsi keanggotaan terbaik untuk variabel input *Word Alignment* adalah segitiga dengan nilai linguistik **Rendah** terletak pada interval [0..

13.64]. Sementara nilai linguistik **Sedang** terletak pada interval [0.. 100]. Dan nilai linguistik **Tinggi** terletak pada interval [13.64.. 100].

7. Aturan fuzzy terbaik untuk sistem identifikasi plagiat yang dihasilkan melalui proses evolusi adalah sebanyak 7 aturan.

5.2 Saran

Beberapa saran yang bisa diberikan antara lain :

1. Bentuk fungsi keanggotaan sistem fuzzy dapat ditambah jenisnya seperti dengan fungsi keanggotaan gaussian, dan lain sebagainya.
2. Jumlah nilai linguistik sistem fuzzy didisain lebih dinamis sehingga dengan kedinamisan tersebut dimungkinkan untuk mengatasi keterbatasan sistem yang telah dibangun yang hanya terdiri dari 3 nilai linguistik. Dimana dengan 3 nilai linguistik tersebut, jumlah aturan fuzzy yang dibangkitkan maksimum sebanyak 9 aturan. Jika nilai linguistik lebih dinamis diharapkan sistem lebih intuitif dan mendapatkan hasil yang lebih baik.
3. Pendefinisian aturan fuzzy dapat dilakukan dengan menggunakan cara *Multiple Fuzzy Grid Table* sehingga aturan fuzzy yang diciptakan bisa lebih optimal.
4. Untuk pelatihan algoritma genetika, dapat dilakukan pengembangan dan percobaan dengan perubahan pada jenis pindah silang yang digunakan, mekanisme penggantian populasi, dan skema pengkodean yang lain untuk mendapatkan pelatihan yang optimal.

Daftar Pustaka

- [1] Away, Gunaidi A. 2006. *The Shortcut of MATLAB programming*. Bandung: Informatika.
- [2] Klir J, George and Yuan, Bo. 1995. *Fuzzy Sets and Fuzzy Logic : Theory and Applications*. Prentice Hall.
- [3] Konar, Amit. 2000. *Artificial Intelligence and Soft Computing: behavioral and cognitive modelling of the human brain*. Florida: CRC Press LLC.
- [4] Kusumadewi, Sri. 2003. *Artificial Intelligence (Teknik dan Aplikasinya)*. Yogyakarta: Graha Ilmu.
- [5] Kusumadewi, Sri dan Purnomo, Hari. 2004. *Aplikasi Logika Fuzzy Untuk Pendukung Keputusan*. Yogyakarta: Graha Ilmu.
- [6] Mitchell M. Tom. 1997. *Machine Learning*, Singapore: McGraw-Hill International Editions.
- [7] Negnevitsky. 2002. *Hybrid Intelligence Systems: Evolutionary neural networks and fuzzy evolutionary systems*. Pearson Education.
- [8] [Http://id.wikipedia.org/wiki/plagiarisme.html](http://id.wikipedia.org/wiki/plagiarisme.html). Di-download pada tanggal 27 Januari 2007.
- [9] [Http://abgaduh.blogspot.com/2004/12/plagiat.html](http://abgaduh.blogspot.com/2004/12/plagiat.html). Di-download pada tanggal 27 Januari 2007.
- [10] Russel, Stuart and Norvig, Peter. 1995. *Artificial Intelligence: A Modern Approach*. Prentice Hall International, Inc.
- [11] Suyanto. 2005. *Algoritma Genetika dalam MATLAB*. Yogyakarta: ANDI.
- [12] Suyanto. 2007. *Artificial Intelligence, Searching, Reasoning, Planning, and Learning*. Bandung: Informatika.
- [13] Suyanto. 2008. *Evolutionary Computation Komputasi Berbasis "Evolusi" dan "Genetika"*. Bandung: Informatika.
- [14] Suyanto. 2008. *Soft Computing Membangun Mesin Ber-IQ Tinggi*. Bandung: Informatika.
- [15] Sedyono, Agung dan Ku Muhammad, Ku Ruhana. 2006. *Usulan Model*

deteksi Plagiasi Dokumen Pada Sistem Dokumen Tersebar yang Kooperatif-Terstruktur. Prosiding Konferensi Nasional Sistem Informasi 2006. Bandung: Informatika.

- [16] Zadeh, Lotfi A.. 2006. *Soft Computing*. www.cs.berkeley.edu/~zadeh
- [17] Hisao Ishibuchi, Tadahiko Murata, and Mitsuo Gen. 1998. *Performance Evaluation of Fuzzy Rule-Based Classification Systems Obtained by Multi-Objective Genetic Algorithms. Computers Industrial Engineering* Vol. 35, Nos 3-4, pp. 575-578.

