Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK) DOI: 10.25126/jtiik.201961227
Vol. 6, No. 2, April 2019, hlm. 123-128 p-ISSN: 2355-7699
Akreditasi KEMENRISTEKDIKTI, No. 30/E/KPT/2018 e-ISSN: 2528-6579

PENGEMBANGAN SISTEM EKSTRAKSI METADATA ARTIKEL ILMIAH SECARA OTOMATIS

Faisal Rahutomo[1], Dyah Ayu Irawati [2], Muhammad Aisamuddin Eka Pramudita[3]

[1,2,3] Jurusan Teknologi Informasi, Politeknik Negeri Malang
Email: [1]faisal@polinema.ac.id, [2]dyah.ayu@polinema.ac.id, [3]muhaisamuddinep@gmail.com

(Naskah masuk: 11 November 2018, diterima untuk diterbitkan: 18 Desember 2018)

Abstrak

Pengarsipan artikel ilmiah di Jurusan Teknologi Informasi Politeknik Negeri Malang menggunakan *platform Open Journal Systems* (*OJS*). Pengarsipan tersebut melalui tahapan penulisan *metadata* artikel ilmiah yang dilakukan satu-persatu. *Metadata* artikel ilmiah ini berupa judul, penulis, instansi, surel penulis, abstrak, kata kunci, dan daftar pustaka. Diperlukan waktu yang cukup lama untuk memasukkan *metadata* artikel ilmiah dalam *OJS* karena prosedur dalam *OJS* itu sendiri. Untuk itu penelitian ini mengusulkan sebuah sistem tambahan *OJS* yang bisa menyediakan *metadata* artikel ilmiah tersebut secara otomatis. Sistem dibangun menggunakan pendekatan *rule-based text parsing*. Dalam metode tersebut disusun beberapa aturan untuk mengambil teks yang diperlukan oleh isian *metadata OJS* yang mewakili sebuah artikel ilmiah. Artikel ilmiah diunggah ke dalam sistem tambahan tersebut untuk menghasilkan *metadata*-nya secara otomatis. *Metadata* tersebut selanjutnya disimpan dalam format *XML*. Pada sistem *OJS* terdapat perangkat *native XML plugin* yang bisa melakukan *export – import metadata* suatu artikel ilmiah untuk *OJS*. Dari hasil pengujian, sistem bisa memudahkan pengarsipan artikel ilmiah lebih cepat 13 kali dibanding pengisian *metadata* secara manual.

Kata kunci: *artikel ilmiah, metadata, rule-based text parsing, open journal systems*

DEVELOPMENT OF AUTOMATIC SCIENTIFIC ARTICLE METADATA EXTRACTION SYSTEM

Abstract

Department of information technology, State Polytechnic of Malang archives its scientific article with Open Journal Systems (OJS) platform. Archiving in OJS needs to write the scientific article metadata manually through a form. Metadata of this scientific article includes title, author, agency, writer e-mail, abstract, keywords and bibliography. Inserting scientific articles metadata in OJS manually takes quite a long time because of the procedure in OJS itself. Highlighting this problem, this research proposes a text processing add-on system for OJS that able to extract the scientific article's metadata automatically. The system is built with rule-based text parsing method. In this method, the authors composed some rules to obtain the metadata of scientific article. Scientific articles were uploaded into the system to capture the metadata of the scientific article automatically. The metadata was then stored in XML. In OJS add-on there is native XML plugin tool that able to export - import the scientific article metadata for OJS. The experimental results show the system able to facilitate the archiving of scientific articles 13 times faster.

Keywords: s*cientific articles*, m*etadata, rule-based text parsing, open journal systems*

1. PENDAHULUAN

Bertambahnya artikel ilmiah setiap tahunnya adalah sebuah kepastian. Hal ini disebabkan setiap universitas wajib menghasilkan artikel ilmiah, baik bagi mahasiswa untuk syarat kelulusan, dan juga bagi dosen sebagai salah satu luaran penelitiannya. Sehingga, semakin tahun jumlah artikel ilmiah terus bertambah semakin banyak. Artikel ilmiah dalam setiap terbitan bisa dan seringkali memiliki aturan format yang berbeda-beda. Meskipun formatnya tampilannya berbeda tetapi dalam tata aturan penulisan artikel ilmiahnya memiliki kemiripan.

Dewasa ini, pengarsipan artikel ilmiah dapat menggunakan kerangka sistem *Open Journal Systems* (*OJS*) (Project, 2016). Petugas yang mengunggah artikel tersebut harus mengetikkan atau memasukkan informasi *metadata* artikel yang diunggah. Sayangnya pembuatan *metadata* masih dilakukan secara manual. Berdasarkan uraian di atas, maka perlu dibangun sebuah sistem yang dapat memudahkan petugas dalam mengunggah artikel

ilmiah. Petugas hanya perlu mengunggah berkas artikel ilmiah tersebut tanpa perlu memasukkan informasi *metadata* yang ada pada *file* tersebut secara manual. Hal ini dimungkinkan karena *metadata* artikel sudah diekstraksi secara otomatis oleh sistem di dalam penelitian ini.

Pendekatan yang dipilih adalah pendekatan praktis, dengan metode *rule-based text parsing* (Russell & Norvig, 2009)(Yates & Neto, 2008)(Manning, Raghavan and Schütze, 2008). Pola teks yang ada di dalam makalah ilmiah dipetakan ciri-cirinya untuk dibuat aturan ekstraksinya. Informasi penting yang telah diekstraksi kemudian disimpan ke dalam format *XML*, menyesuaikan *add-on OJS* yang telah tersedia. Dengan demikian makalah diharapkan dapat diarsipkan secara lebih cepat tanpa kerepotan berlebih memasukkan metadata artikel satu-persatu.

Makalah ini disusun ke dalam lima bagian. Bagian pertama membahas pendahuluan. Bagian kedua membahas studi literatur yang mendasari penelitian ini. Bagian ketiga membahas metode yang diusulkan pada makalah ini. Bagian keempat membahas pengujian dan pembahasan. Diakhiri oleh bagian kelima yang memaparkan simpulan dari penelitian ini.

2. TINJAUAN PUSTAKA

2.1 *Metadata*

Metadata adalah data tentang data. Sebuah *file*, yang merupakan sebuah data memiliki data lain yang menginformasikan tentang *file* tersebut. *Metadata* adalah informasi terstruktur yang mendeskripsikan, menjelaskan, menemukan, atau setidaknya membuat menjadikan suatu informasi mudah untuk ditemukan kembali, digunakan, atau dikelola. *Metadata* sering disebut sebagai data tentang data atau informasi tentang informasi. *Metadata* ini mengandung informasi mengenai isi dari suatu data yang dipakai untuk keperluan manajemen *file* atau data itu nantinya dalam suatu basis data (NISO, 2004).

Metadata untuk artikel adalah informasi mengenai artikel. Misalnya data pengarang, judul artikel, tahun, tanggal publikasi. Dapat pula merupakan bahan deskriptif seperti kata kunci dan abstrak. Sehingga, untuk memudahkan pengarsipan dan pencarian kembali sebuah artikel ilmiah dibutuhkan sebuah *metadata* dari artikel ilmiah itu sendiri.

2.2 *Text Parsing*

Text parsing adalah teknik komputer membaca teks dan memaknainya. Sebuah dokumen dikenali sebagai kumpulan kata-kata atau teks biasa tanpa format tertentu. Kemudian teks di dalam dokumen tersebut dipecah menjadi bagian-bagian kecil yang bisa dimengerti. Dalam kasus *kompiler* pemahaman komputer terhadap bagian-bagian kecil teks tersebut dimanfaatkan lebih lanjut ke dalam tahap kompilasi. Sedangkan pada kasus ekstraksi informasi, pemahaman bagian-bagian kecil teks tersebut dipahami sebagai informasi yang berhasil diekstrak dari sebuah dokumen. Agar komputer dapat memahaminya bagian kecil teks yang dimaksud, diperlukan ciri-ciri dan tanda-tanda khusus atau aturan sintaks tertentu di dalam dokumen (Manning & Schütze, 1999).

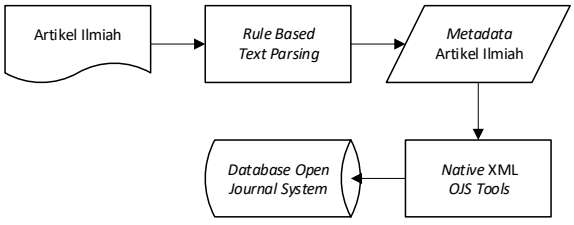
2.3 *Rule-Based Expert System*

Rule-based adalah salah satu pendekatan atau algoritma dalam ilmu teknologi informasi yang bisa digunakan untuk berbagai permasalahan. *Rule-based* sesuai untuk semua bidang yang mana area masalahnya dapat ditulis dalam bentuk pernyataan aturan *if-then*. Rangkaian aturan if-then ini bisa disusun berantai hingga menghubungkan antara anteseden dan konsekuensi. Relasi logika atau, tidak, atau dan dapat dikombinasikan di dalam aturan *if-then* yang dibangun (Russell & Norvig, 2009).

2.4 *OJS*

OJS adalah sistem yang digunakan untuk mengelola artikel ilmiah. Ia dirancang untuk memfasilitasi pengembangan akses terbuka publikasi ilmiah dengan *peer-review*. *OJS* menyediakan infrastruktur teknis tidak hanya untuk presentasi artikel jurnal secara daring, namun juga keseluruhan alur kerja manajemen editorial. Termasuk di dalamnya kiriman artikel, beberapa putaran *peer-review*, dan pengindeksan. Dalam situs resminya (Project, 2016) disebutkan, *OJS* adalah perangkat lunak terbuka yang tersedia secara bebas untuk jurnal di seluruh dunia. *OJS* membuat sistem akses terbuka agar ada lebih banyak jurnal terbuka. Karena, akses yang terbuka dapat meningkatkan pembaca jurnal serta kontribusinya terhadap kebaikan publik dalam skala global.

3. EKSTRAKSI *METADATA*



Gambar 1. *Diagram alir sistem*

Pada penelitian ini *metadata* yang diekstrak adalah *metadata* dari suatu artikel ilmiah. Gambar 1 mendeskripsikan langkah-langkah yang dilakukan di dalam usulan penelitian ini. Pertama, artikel ilmiah diubah menjadi teks biasa tanpa format. Selanjutnya penyusunan serangkaian aturan *if-then* atas ciri-ciri yang dapat dikenal di dalam makalah ilmiah atas

informasi-informasi apa saja yang ingin diekstrak dari dokumen. Selanjutnya kumpulan aturan ini diimplementasikan dalam bentuk *regular expression*. Luaran dari aturan-aturan ini adalah sekumpulan informasi yang kemudian disusun dan disesuaikan ke dalam bentuk standar *native XML plugin*. Untuk itu percobaan *export XML* dilakukan terlebih dahulu atas sebuah makalah yang telah diunggah manual untuk mengenali struktur *XML* yang dimaksud. Setelah itu sistem diintegrasikan untuk menjadi sebuah perangkat yang berguna.

3.1 Pengumpulan Data

Pada tahap ini pengumpulkan data dan informasi yang diperlukan dilakukan untuk melakukan ekstraksi *metadata* artikel ilmiah. Artikel ilmiah yang dimaksud di dalam makalah ini berasal dari laman jurnal Politeknik Negeri Malang (Polinema), lebih spesifik di dalam Jurnal Informatika Polinema (JIP) (Malang, 2018). Artikel ilmiah yang digunakan adalah publikasi dari tahun 2014 hingga 2017. Penelitian ini masih membatasi karena satu format artikel ilmiah. Dengan demikian aturan-aturan di dalam algoritma *rule-based text parsing* dibatasi pada satu format artikel ilmiah JIP. Algoritma rule-based yang dinamik dapat dikembangkan lebih lanjut sebagai pengembangan makalah ini. Sistem yang dibangun dibatasi hanya menerima masukkan berupa berkas berformat pdf.

3.2 Pengembangan Sistem

Pada sistem ini proses pengembangan dilakukan dengan menggunakan bahasa pemrograman PHP. Untuk melakukan pembacaan berkas pdf digunakan komponen PdfParser yang dikembangkan oleh Malot (Malot, 2018). Bagian lainnya dari program dibangun dengan PHP. Pengembangan aplikasi ini diikuti dengan pengujian yang bertujuan untuk mengecek fungsi setiap fitur yang terdapat pada sistem agar setiap fitur yang ada bekerja sesuai perancangan. Proses pengujian pada sistem yang dikembangkan menggunakan web browser dan dilakukan dengan menguji fungsionalitas dari setiap fitur yang tersedia pada sistem. Pengujian ini dilakukan setiap saat setelah satu fitur selesai dibuat. Proses ini dilakukan hingga setiap fitur bekerja sesuai dengan yang direncanakan.

Antarmuka dari halaman awal ditunjukkan oleh Gambar 2. Bagian atas halaman awal ini berisi *form upload* berkas. Bagian tengahnya berisi tabel yang berisi daftar artikel ilmiah yang sebelumnya sudah pernah diunggah. Kemudian terdapat beberapa tombol untuk menghapus, melihat detail dan mengunduh *file XML* yang berisi *metadata* dari artikel ilmiah. Antarmuka dari detail *metadata* menampilkan detail *metadata* dari artikel ilmiah yang sudah diunggah berupa *popup* seperti yang ditunjukkan pada Gambar 3. Pengguna dapat mengevaluasinya sebelum digunakan lebih lanjut. Apabila tombol *extract* ditekan, berkas *XML* dapat diunduh untuk digunakan lebih lanjut.

Gambar 2. Halaman awal

Gambar 3. Detail *metadata*

3.3 Implementasi *Rule Based Text Parsing*

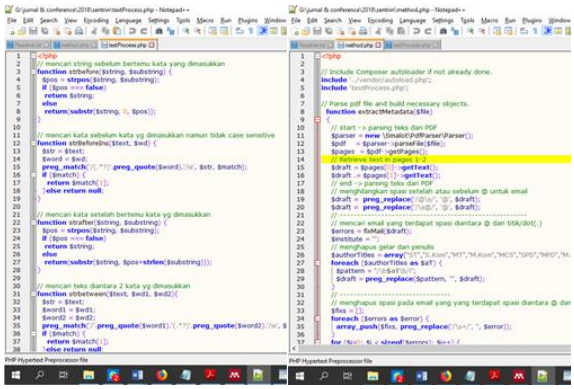
Penelitian ini merancang alur kerja sistem berdasarkan penelitian Zamzami dkk (Zamzami, dkk, 2016), yang juga menggunakan algoritme r*ule-based text parsing.* Pada penelitian tersebut dilakukan percobaan penyusunan *WordNet* Bahasa Indonesia yang diambil dari sumber Kamus Tesaurus Bahasa Indonesia (Bahasa, 2008).

Tabel 1. Tabel Ciri-Ciri yang Dikenali

ciri-ciri	informasi
Sebelum bertemu kata / kalimat dengan huruf besar semua atau diawali dengan kata Volume dan diakhiri dengan angka sebelum bertemu dengan kata.	*Header* (Volume, Edisi dan Tahun)
Kata / kalimat dengan huruf besar semua.	Judul
Setelah kata / kalimat dengan huruf besar semua, sebelum bertemu angka.	Nama
Setelah angka yang dibatasi koma.	Instansi
Kata yang terdapat karakter @, dibatasi dengan koma (,).	Email
Setelah kata Abstrak sebelum bertemu kata kunci.	Abstrak
Setelah kata kunci, dibatasi koma, sebelum pendahuluan.	Kata kunci
Dimulai setelah kata (daftar pustaka:). Kemudian referensi tersebut dikategorikan lagi menurut nama penulis, tahun terbit, judul, dan sumber yang dipisahkan dengan karakter titik (.).	Referensi

Setelah data yang berupa artikel tersebut didapatkan, maka langkah selanjutnya adalah mengidentifikasi ciri-ciri penting yang dapat dikenali dari dokumen untuk mengambil informasi-informasi tertentunya saja. Tabel 1 menjelaskan ciri-ciri yang dikenali di dalam dokumen, terkait dengan

informasi-informasi yang diinginkan untuk diambil di dalam metode ini.



Gambar 4. Implementasi program PHP

Gambar 4 menunjukkan *screenshoot* sebagian implementasi *rule-based* di dalam kode PHP. Kode di dalam penelitian ini akan dibuat akses terbuka melalui lama Github setelah artikel ini terbit.

4. PENGUJIAN DAN PEMBAHASAN

Dalam tahap ini dilakukan uji coba aplikasi yang dibuat. Pengujian aplikasi yang dilakukan meliputi ekstraksi aplikasi apakah sudah sesuai dengan yang diharapkan. Hasil pengujian menunjukkan pada akhirnya semua fitur aplikasi dapat berfungsi sebagaimana yang direncanakan.

Setelah setiap fitur berfungsi dengan benar, tahap selanjutnya adalah tahap pengujian performa sistem. Tujuannya untuk evaluasi apakah usulan di dalam penelitian ini memiliki performa yang lebih baik dibandingkan tanpanya. Pada tahap ini, akan dilakukan beberapa metode pengujian, yaitu pengukuran akurasi dan waktu sistem dalam melakukan ekstraksi. Pengukuran akurasi dilakukan untuk mengukur tingkat kesesuaian ekstraksi metadata oleh sistem dengan ekstraksi metadata yang dilakukan secara manual. Pengujian waktu dilakukan untuk mengukur tingkat kecepatan ekstraksi metadata artikel ilmiah oleh sistem dibandingkan dengan ekstraksi yang dilakukan secara manual untuk dimasukkan ke dalam OJS.

4.1 Akurasi Sistem

Pengujian ini dilakukan dengan mengambil *file* uji coba sebanyak 10% dari setiap Volume Jurnal Informatika Polinema. Pengujian ini dilakukan pada artikel yang telah terbit, yang telah ditulis dengan format tata tulis yang sesuai dengan *template* yang disediakan. Tabel 2 mencatat hasil percobaan yang dilakukan. Dari data tersebut maka dapat dihitung tingkat akurasi sistem terlihat pada perhitungan dengan rumus seperti di bawah ini:

$$\text{Akurasi}=\frac{7\times15-5}{7\times15}\times100\%=95\% \quad (1)$$

Dengan demikian dapat diketahui bahwa sistem dapat bekerja dengan baik apabila dari *file* pdf dari artikel ilmiah bisa di-*parsing* dengan baik. Selain itu format aturan penulisan dalam jurnal juga sangat berpengaruh dalam kinerja sistem ini, karena menggunakan metode r*ule-based text-parsing* yang statik, belum dinamik.

Tabel 2. Akurasi Sistem

NO	*Header*	Judul	Nama	Instansi	Email	Abstrak	Kata Kunci
1	✓	✓	✓	✓	✓	✓	✓
2	✓	✓	✓	✓	✓	✓	✓
3	✓	✓	✓	✓	✓	✓	✓
4	✓	✓	✓	✓	✓	✓	✓
5	✓	✓	✓	✓	✓	✓	✓
6	✓	✓	✓	✓	-	✓	✓
7	✓	✓	✓	✓	✓	✓	✓
8	✓	✓	✓	✓	✓	✓	✓
9	✓	✓	✓	✓	✓	✓	✓
10	✓	✓	✓	✓	✓	✓	✓
11	✓	✓	✓	✓	✓	✓	✓
12	✓	✓	✓	✓	✓	✓	✓
13	✓	✓	✓	✓	✓	✓	✓
14	✓	-	-	✓	✓	✓	✓
15	✓	-	-	✓	✓	✓	✓

4.2 Waktu Kinerja Sistem

Pengujian ini dilakukan dengan meminta 3 orang pengguna untuk memasukkan 10 data artikel ilmiah pada OJS secara manual dan dengan sistem yang dibuat. Kemudian dibandingkan lebih cepat mana memasukkan data artikel ilmiah dengan sistem atau secara manual. Tabel 3 menunjukkan hasil dari pengujian tersebut.

Tabel 3. Pengujian Waktu

No.	Waktu	
	Manual	**Sistem**
1	1 menit 1 detik	14 menit 53 detik
2	1 menit 35 detik	20 menit 29 detik
3	1 menit 20 detik	16 menit 41 detik
Rata-rata	78,67 detik	1041 detik

Dari data tersebut dapat diproses untuk mengukur efisiensi waktu dalam mengunggah artikel ilmiah menggunakan sistem dibandingkan dengan secara manual pada OJS. Perhitungan efisiensi waktu dilakukan seperti pada rumus di bawah ini:

$$\text{Perbandingan Waktu} = \frac{Manual\ (detik)}{Sistem\ (detik)} \quad (2)$$

Dengan rumus tersebut dapat dihitung perbandingan waktu menggunakan sistem adalah 13,23 kali lebih cepat dibanding memasukkan data secara manual.

5. KESIMPULAN

Berdasarkan hasil pengujian dapat disimpulkan bahwa Sistem Ekstraksi *Metadata* Artikel Ilmiah dapat dibangun dengan menggunakan metode *rule-based text parsing.* Sistem yang dibangun berhasil mengekstraki *metadata* dari artikel ilmiah JIP dengan baik, memiliki nilai galat kurang dari 10%. Galat yang ada dimungkinkan karena kesalahan pembacaan karakter berkas pdf ke teks oleh komponen PdfParser. Kemungkinan lainnya dikarenakan kesalahan penulisan artikel, tidak sesuai *template*. Selain itu sistem ekstraksi *metadata* artikel ilmiah ini lebih efisien dalam memasukkan data artikel ilmiah pada *OJS*. Pengujian mencatat pendekatan dengan sistem 13,23 kali lebih cepat dibanding masukan secara manual.

Saran yang dapat diberikan dari hasil penelitian untuk pengembangan sistem ini ke depan yaitu ekstraksi dengan metode yang lebih cerdas dan dinamik, yang dapat mengenali format artikel ilmiah dengan *template* yang berbeda-beda.

DAFTAR PUSTAKA

BAEZA-YATES, R. AND RIBEIRO-NETO, B., 2008. *Modern Information Retrieval: The Concepts and Technology Behind Search*. 2nd edn. USA: Addison-Wesley Publishing Company.

BAHASA, P., 2008. *Kamus Tesaurus Bahasa Indonesia*. Departemen Pendidikan Nasional.

MALANG, J. T. I. P. N., 2018. *Jurnal Informatika Polinema (JIP)*. Available at: http://jip.polinema.ac.id/ojs3/index.php/jip.

MALOT, S., 2018. *PDF Parser*. Available at: https://pdfparser.org/.

MANNING, C. D., RAGHAVAN, P. AND SCHÜTZE, H., 2008. *Introduction to Information Retrieval*. New York, NY, USA: Cambridge University Press.

MANNING, C. D. AND SCHÜTZE, H., 1999. *Foundations of Statistical Natural Language Processing*. Cambridge, MA, USA: MIT Press.

NISO, 2004. *Understanding metadata*. 4733 Bethesda Avenue, Suite 300, Bethesda, MD 20814 USA: NISO. Available at: http://www.niso.org/publications/press/UnderstandingMetadata.pdf.

PROJECT, P. K., 2016. *Open Journal Systems*. Available at: https://pkp.sfu.ca/ojs/.

RUSSELL, S. AND NORVIG, P., 2009. *Artificial Intelligence: A Modern Approach*. 3rd edn. Upper Saddle River, NJ, USA: Prentice Hall Press.

ZAMZAMI, D., RAHUTOMO, F. AND PUSPITASARI, D., 2016. 'Aplikasi Wordnet Indonesia Berdasarkan Kamus Thesaurus Bahasa Indonesia menggunakan Algoritma Rule Based Text Parsing', in *Seminar Informatika Aplikatif Polinema*. Jurusan Teknologi Informasi Politeknik Negeri Malang.

Halaman ini sengaja dikosongkan