

LAPORAN PENELITIAN



KLASTERING DOKUMEN BERITA DARI WEB MENGUNAKAN ALGORITMA SINGLE PASS CLUSTERING

Oleh :

1. Hery Februriyanti, ST., M.Cs / YS.2.01.01.035 (Ketua)
2. Drs. Eri Zuliarso, M.Kom / YS.2.93.01.093 (Anggota)
3. Rina Anwaristyati, S.Kom / Y.3.96.12.066 (Anggota)
4. Gegit Hendriatmono / 07.01.53.0191 (Anggota)
5. Jupriyanto / 07.01.53.0118 (Anggota)

FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS STIKUBANK (UNISBANK) SEMARANG
SEPTEMBER 2011

HALAMAN PENGESAHAN LAPORAN PENELITIAN

1. a. Judul Pelatihan : Klastering Dokumen Berita Dari Web
Menggunakan Algoritma Single Pass Clustering
b. Bidang Ilmu : Komputer (Sistem Informasi)
-
2. Ketua Peneliti
a. Nama Lengkap : Herny Februariyanti, ST., M.Cs
b. Jenis Kelamin : Perempuan
c. NIP : YS.2.01.01.035
e. Disiplin Ilmu : Ilmu Komputer
f. Pangkat / Golongan : Penata Muda / IIC
g. Jabatan : Lektor
h. Fakultas / Jurusan : Teknologi Informasi / Sistem Informasi
i. Alamat : Jl. Trilomba Juang No. 1 Semarang
j. Telp / Faks / Email : 024-8311668/024-8443240/
info@unisbank.ac.id
k. Alamat Rumah : Jl. Kendeng V/12, Semarang
h. Telp / Faks / Email : 08156545909/ - / herny@unisbank.ac.id
-
3. Jumlah Anggota Peneliti : 4 orang
Nama Anggota : Drs. Eri Zuliarso, M.Kom / YS.2.93.01.093
Nama Anggota : Rina Anwaristyati, S.Kom / Y.3.96.12.066
Nama Anggota : Gegit Hendriatmono / 07.01.53.0191
Nama Anggota : Jupriyanto / 07.01.53. 0118
-
4. Lokasi Penelitian : Laboratorium Komputer Unisbank Semarang
-
5. Jangka waktu Penelitian : 3 (tiga) bulan, 10 Juli 2011 - 30 September 2011
6. Jumlah biaya yang diusulkan : Rp. 3.000.000,00
-

Mengetahui,
Dekan Fakultas Teknologi Informasi

Semarang, 30 September 2011
Ketua Pelaksana,

(Dwi Agus Diartono, S.Kom., M.Kom)
NIY: Y.2.90.03.054

(Herny Februariyanti, ST., M.Cs)
NIY: YS.2.01.01.035

Menyetujui,
Ketua LPPM Unisbank

(Dr. Dra. Lie Liana, M.MSI)
NIY. Y.2.92.07.085

KATA PENGANTAR

Bismillahirrohmanirrohim. Puji syukur kehadiran Allah SWT atas limpahan rahmat-Nya sehingga laporan penelitian yang berjudul "Klastering Dokumen Berita dari Web Menggunakan Algoritma Single Pass Clustering" ini dapat diselesaikan.

Penulisan penelitian ini dapat terselesaikan oleh penulis, tentunya tidak lepas dari berbagai pihak yang memberikan bimbingan dan dorongan serta menyumbangkan tenaga, pikiran, dan perhatian. Oleh karena itu, pada kesempatan ini penulis menyampaikan terima kasih yang sebesar – besarnya kepada :

1. Bapak Dr. Bambang Suko Priyono, SE., MM., selaku Rektor Universitas Stikubank (Unisbank) Semarang.
2. Ibu Dr. Dra., Lie Liana, M.MSI, selaku Ketua Lembaga Penelitian dan Pengabdian Masyarakat (LPPM) Universitas Stikubank (Unisbank) Semarang.
3. Bapak Dwi Agus Diartono, M.Kom., selaku Dekan Fakultas Teknologi Informasi Universitas Stikubank (Unisbank) Semarang.
4. Rekan-rekan dosen yang telah memberikan masukan-masukkan untuk perbaikan dan kesempurnaan penulisan laporan ini.

Semoga penu

lisan laporan penelitian ini dapat bermanfaat dan menambah ilmu bagi semua serta dapat mendukung kemajuan ilmu pengetahuan khususnya di bidang Teknologi Informasi.

Semarang, 30 September 2011

Penulis

ABSTRAK

KLASTERING DOKUMEN BERITA DARI WEB MENGGUNAKAN ALGORITMA SINGLE PASS CLUSTERING

Herny Februariyanti, ST., M.Cs, Drs. Eri Zuliarso, M.Kom.,
Rina Anwaristyati, S.Kom, Gegit Hendriatmono, Jupriyanto

Fakultas Teknologi Informasi, Program Sistem Informasi
Universitas Stikubank (UNISBANK) Semarang

Usaha untuk memperoleh informasi secara digital telah banyak dilakukan dan perkembangannya sangat pesat seiring dengan perkembangan teknologi komputer. Dalam proses penelusuran informasi melalui internet sering diperoleh informasi yang sangat banyak, tetapi sebagian besar diantaranya adalah informasi yang tidak dibutuhkan. Oleh karena itu, dari sudut pandang temu kembali informasi (*information retrieval*), semakin banyaknya informasi yang tersedia di internet justru semakin mempersulit untuk menemukan kembali informasi yang relevan, yaitu informasi yang sesuai dengan kebutuhan.

Dalam suatu sistem temu kembali informasi, kemampuan untuk menemukan informasi yang tersedia diukur dengan *recall* dan kemampuan untuk menemukan informasi yang relevan diukur dengan ketelitian, maka proses penelusuran dalam situasi seperti tersebut di atas akan menghasilkan *recall* yang tinggi tetapi ketelitian rendah.

Penelitian ini berusaha untuk mengkluster dokumen dengan menggunakan Algoritma *Single Pass Clustering*. Klastering ini ditekankan untuk dokumen

berbahasa Indonesia. Keterkaitan antar dokumen diukur berdasarkan kemiripan antar dokumen (*similarity*).

Algoritma ini diuji coba dengan menggunakan dokumen berita arsip berita online Kompas <http://www.kompas.com/archive> dalam format HTML. Hasil uji coba menunjukkan bahwa algoritma ini dapat diaplikasikan untuk pengelompokan dokumen berbahasa Indonesia. Pemilihan kata kunci yang tepat akan meningkatkan kualitas temu kembali informasi (*information retrieval*) pada dokumen. Hasil penelitian di dapatkan hasil terbaik pada threshold 0.2 dengan hasil recall sebesar 0.33 dan precision 0.37 .

Kata kunci : *information retrieval, simmilaritas, single pass clustering, recall, precision*