KATALOG BPS: 1202031
ISSN: 2086–4132

JURNAL APLIKASI STATISTIKA & KOMPUTASI STATISTIK

TAHUN 6, VOLUME 2, DESEMBER 2014

Kajian Penghitungan Nilai Tukar Petani Tanaman Pangan (NTPP) di Jawa, Bali, dan Nusa Tenggara Tahun 2011 – 2013
EKARIA dan ATIKA NASHIRAH HASYYATI

Metode *C-Means Cluster* dan *Fuzzy C-Means Cluster* pada Kasus Pengelompokan Desa Menurut Status Ketertinggalan (Studi di Kota Metro dan Kabupaten Lampung Timur)
SUKIM

Pengaruh *Foreign Direct Investment* (*FDI*) terhadap Pertumbuhan Ekonomi 10 Negara *ASEAN*
AISYAH FITRI YUNIASIH

Daya Saing dan Variabel yang Memengaruhi Ekspor Batubara Indonesia di Delapan Negara Tujuan Ekspor Tahun 2002-2012
HARIANTO SARDY PURBA dan FITRI KARTIASIH

Framework untuk Mendeteksi Pemalsuan Data pada *Mobile Survey*
IBNU SANTOSO

Pengembangan Sistem *Web Crawler* Sebagai Sarana Riset Media Secara Otomatis (Studi di Subdit Neraca Rumah Tangga dan Institusi Nirlaba)
ENGGELIN GIACINTA WONGKAR dan YUNARSO ANANG SULISTIADI

UNIT PENELITIAN DAN PENGABDIAN KEPADA MASYARAKAT
SEKOLAH TINGGI ILMU STATISTIK
(UPPM-STIS)

TAHUN 6, VOLUME 2, DESEMBER 2014

KATALOG BPS: 1202031
ISSN: 2086-4132

JURNAL APLIKASI STATISTIKA & KOMPUTASI STATISTIK

Kajian Penghitungan Nilai Tukar Petani Tanaman Pangan (NTPP) di Jawa, Bali, dan Nusa Tenggara Tahun 2011 – 2013 1-18
EKARIA dan ATIKA NASHIRAH HASYYATI

Metode *C-Means Cluster* dan *Fuzzy C-Means Cluster* pada Kasus Pengelompokan Desa Menurut Status Ketertinggalan (Studi di Kota Metro dan Kabupaten Lampung Timur) 19-51
SUKIM

Pengaruh *Foreign Direct Investment* (*FDI*) terhadap Pertumbuhan Ekonomi 10 Negara *ASEAN* 52-68
AISYAH FITRI YUNIASIH

Daya Saing dan Variabel yang Memengaruhi Ekspor Batubara Indonesia di Delapan Negara Tujuan Ekspor Tahun 2002-2012 69-93
HARIANTO SARDY PURBA dan FITRI KARTIASIH

Framework untuk Mendeteksi Pemalsuan Data pada *Mobile Survey* 94-114
IBNU SANTOSO

Pengembangan Sistem *Web Crawler* Sebagai Sarana Riset Media Secara Otomatis (Studi di Subdit Neraca Rumah Tangga dan Institusi Nirlaba) 115-139
ENGGELIN GIACINTA WONGKAR dan YUNARSO ANANG SULISTIADI

PENGEMBANGAN SISTEM WEB CRAWLER SEBAGAI SARANA RISET MEDIA SECARA OTOMATIS

(Studi di Subdit Neraca Rumah Tangga dan Institusi Nirlaba)

Enggelin Giacinta Wongkar
Staf Badan Pusat Statistik

Yunarso Anang Sulistiadi
Dosen Sekolah Tinggi Ilmu Statistik

Abstrak

With the vast development of data to become informations on the Internet, everything online seems to explode at a rapid rate. These informations, including online news which is created as a complement to the original printed media, has even overtaken the latter. Subdirectorate of Household National Account and Non-profit Institution of Statistics Indonesia is in charge for the work of media research. In the process of media research, time and human resources are two important elements but yet having problem of ineffective and inefficient process. This study aimed to overcome that problem by developing a web crawler system that could do summarization automatically from online news sites (currently from Bisnis and Kontan) with output in Microsoft Word format file and minimizing number of similar news. This system is developed using several techniques in information technologies such as crawling and wrapping method and cosine similarity method to minimalize similar news. The result shows the process of media research by using this system much more effective and efficient.

Keywords: web crawler, wrapper, cosine similarity, information extraction

I. PENDAHULUAN

Badan Pusat Statistik (BPS) merupakan Lembaga Pemerintah Non-Kementrian yang bertanggung jawab langsung kepada Presiden. Berdasarkan UU Nomor 16 Tahun 1997 tentang Statistik, salah satu peran yang harus dijalankan oleh BPS adalah menyediakan kebutuhan data bagi pemerintah dan masyarakat. Di antara media yang digunakan oleh BPS untuk menyampaikan atau menyebarkan data statistik adalah *press release*. Satu di antara sekian

agenda *press release* BPS adalah penyajian angka-angka ekonomi kepada publik setiap tiga bulan sekali / triwulanan yang dilakukan oleh Direktorat Neraca Pengeluaran, khususnya Subdirektorat Neraca Rumah Tangga dan Institusi Nirlaba.

Angka-angka ekonomi yang disajikan pada saat *press release* tentu saja harus dapat merepresentasikan fenomena-fenomena yang jelas dan nyata yang terjadi di Indonesia pada saat itu. Misalnya, hasil suatu penelitian menyebutkan bahwa nilai tukar petani di Propinsi Jawa Timur pada suatu tahun tertentu mengalami penurunan. Penurunan ini harusnya merupakan dampak dari suatu fenomena yang terjadi kepada petani di Propinsi Jawa Timur. Mungkin pada tahun tersebut, petani-petani di Jawa Timur mengalami gagal panen yang berakibat pada menurunnya kesejahteraan petani sehingga berkontribusi pada penurunan nilai tukar petani. Fenomena-fenomena seperti itulah yang akan dicari dan diteliti oleh Subdirektorat Neraca Rumah Tangga dan Institusi Nirlaba. Untuk mencari dan meneliti fenomena-fenomena tersebut, Subdirektorat Neraca Rumah Tangga dan Institusi Nirlaba menggunakan informasi dari situs-situs berita yang ada, baik di media cetak maupun di Internet (situs berita *online*). Proses pengumpulan dan penginterpretasian data ini yang kemudian disebut sebagai riset media.

Seiring berkembangnya teknologi informasi, Internet menjadi ajang komunikasi yang sangat cepat dan efektif sehingga telah menyimpang jauh dari misi awalnya (Nurudin, 2007). Internet telah tumbuh menjadi kebutuhan utama manusia sebagai alat informasi dan komunikasi yang tidak dapat diabaikan (Safitri, 2010). Hal ini berdampak pada makin banyaknya konten *online* di Internet, baik konten yang bersifat hiburan maupun yang bersifat pendidikan dan pengetahuan. Situs-situs berita ekonomi *online* seperti Bisnis, Kontan, Kompas, maupun Antara juga tidak luput dari dampak ini. Konten berita ekonomi pada situs-situs berita *online* tersebut semakin meningkat seiring waktu. Akibatnya, proses pembacaan dan perangkuman berita-berita ekonomi dari situs berita ekonomi *online* akan semakin memakan waktu. Oleh karenanya, dibutuhkan kegiatan perangkuman berita yang lebih efektif dan efisien dalam hal waktu dan tenaga.

Tujuan umum dari penelitian ini adalah untuk mencari metode dan solusi agar waktu dan tenaga yang digunakan pada proses pembacaan dan perangkuman berita-berita ekonomi *online* dari Internet dapat diminimalisasi. Dari sisi waktu, diharapkan dapat mengurangi waktu pembacaan dan perangkuman berita sebesar lebih dari 50% dari total waktu yang dibutuhkan untuk melakukan perangkuman berita. Dari sisi tenaga, diharapkan dapat mengurangi keterlibatan pengguna atau aktivitas yang dilakukan pengguna untuk membaca dan merangkum berita sekitar sepertiga dari total aktivitas yang dilakukan.

Secara khusus, tujuan dari penelitian ini adalah sebagai berikut.

1. Membantu pengguna (dalam hal ini *subject matter*) untuk membuat rangkuman berita dalam format *Microsoft Word* setiap hari untuk berbagai kepentingan, antara lain untuk keperluan pembuatan riset media, agar dapat mengetahui fenomena ekonomi secara cepat, serta membantu orang-orang di subdirektorat untuk dapat membaca bermacam-macam berita yang berkaitan dengan ekonomi dalam waktu singkat.
2. Membantu pengguna untuk membaca berita-berita *online* secara terintegrasi pada satu tempat.
3. Membantu pengguna untuk dapat menemukan berita yang menjadi kehendaknya dengan memberikan fasilitas pengkategorian berita berdasarkan kata kunci yang ditentukan oleh pengguna serta memberikan fasilitas pencarian berita berdasarkan kategori, waktu, dan sumber berita.

Manfaat dari penelitian ini adalah meningkatkan efisiensi dan efektifitas pekerjaan subdirektorat Neraca Rumah Tangga dan Institusi Nirlaba dalam mencari, merangkum, dan menganalisis berita-berita yang dibutuhkan untuk dapat menangkap fenomena yang mendukung data-data ekonomi yang disajikan.

Adapun batasan dari penelitian yang dilakukan penulis ini antara lain:

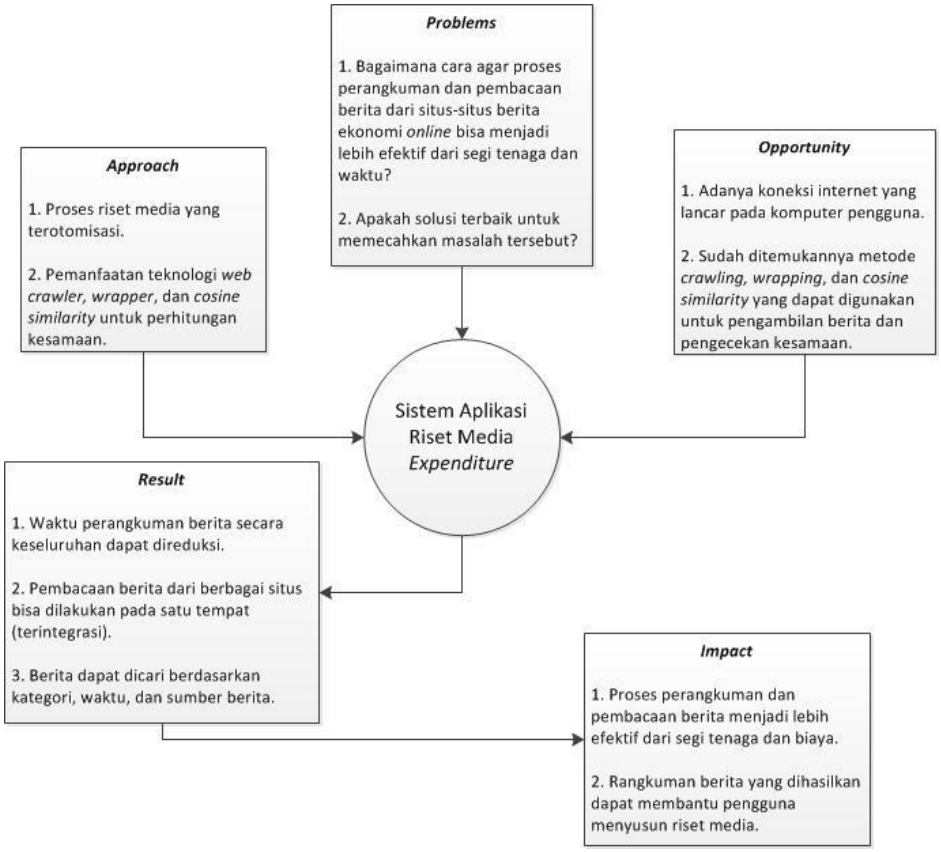
1. Situs-situs berita *online* yang menjadi sumber pengambilan berita dari sistem aplikasi ini hanya terbatas pada situs-situs berita *online* permintaan *subject matter*, dalam penelitian kali ini situs berita ekonomi *online* yang digunakan ada dua, yaitu Bisnis (http://finansial.bisnis.com/) dan Kontan (http://keuangan.kontan.co.id/).
2. Meta-data dan elemen yang diambil dari berita-berita *online* adalah judul, headline, isi berita, waktu berita pada situs bersangkutan, lokasi pengambilan berita, penulis (jika tercantum), editor, dan sumber berita.
3. Dalam perangkuman berita dengan mengambil kalimat yang memiliki angka saja, tidak diperhatikan arti semantiknya.
4. Dalam pengecekan kesamaan dari dua berita, jika ada dua berita yang sama, maka yang akan diambil dan diproses lebih lanjut adalah salah satu dari kedua berita tersebut sesuai dengan algoritma yang telah ditentukan penulis.
5. Waktu yang digunakan sebagai atribut berita adalah waktu saat diambilnya berita, bukan waktu yang tertera pada berita dalam situs yang bersangkutan.

II. TINJAUAN PUSTAKA

1. Kerangka Pikir

Sebagaimana terlihat pada Gambar 1, proses pembangunan Sistem Aplikasi Riset Media *Expenditure* ini terdiri atas 5 komponen, yaitu *problems*, *approach*, *opportunity*, *result*, dan *impact*. Kelima komponen ini dapat dijelaskan sebagai berikut.

- *Problems*: Bagaimana cara agar proses perangkuman dan pembacaan berita dari situs-situs berita ekonomi online bisa menjadi lebih efektif dari segi tenaga dan waktu, serta apakah solusi terbaik untuk memecahkan masalah tersebut.
- *Opportunity*: Adanya koneksi internet yang lancar pada komputer pengguna serta telah ditemukannya metode *crawling*, *wrapping*, dan *cosine similarity* untuk pengambilan berita dan pengecekan kesamaan.
- *Approach*: Proses riset media yang terotomisasi serta pemanfaatan teknologi *web crawler*, *wrapper*, dan *cosine similarity* untuk perhitungan kesamaan.
-



Gambar 1. Kerangka Pikir Sistem Aplikasi Riset Media *Expenditure*

- *Result*: Waktu perangkuman berita secara secara keseluruhan dapat direduksi, pembacaan berita dari berbagai situs bisa dilakukan pada satu tempat (terintegrasi), serta berita yang dapat dicari berdasarkan kategori, waktu, dan sumber berita.

- *Impact*: Proses perangkuman dan pembacaan berita menjadi lebih efektif dari segi tenaga dan biaya, serta rangkuman berita yang dihasilkan dapat membantu pengguna membuat riset media.

2. Penelitian Terkait

Penelitian terdahulu yang berhubungan dengan penelitian ini adalah penelitian yang dilakukan I Gede Arya Agus Yogantara dalam skripsinya yang berjudul "Pengembangan Sistem Aplikasi Pengelolaan *Resume* Berita Media Cetak dan *Online* secara Terpadu" (2010). Pada penelitian tersebut, pengguna mengetikkan sendiri berita yang ingin dikelola dari media cetak maupun media *online*. Perbedaan penelitian tersebut dengan penelitian ini adalah pada penelitian ini, berita yang ada pada situs berita *online* diambil dan dikelola secara otomatis. Selain itu, penelitian ini hanya terbatas pada media *online*.

III. METODOLOGI

Tahapan pembangunan yang dilakukan penulis dalam penelitian ini adalah *Design Oriented Research*, yaitu membuat sebuah artifak baru yang dapat memudahkan pengguna dalam memanfaatkan hasil implementasi dari penelitian ini. Adapun langkah-langkah pembangunan artifak tersebut adalah sebagai berikut.

1. *Awareness of Problem*, yaitu menemukan masalah utama pada sistem berjalan, dalam hal ini pada Subdirektorat Neraca Rumah Tangga dan Institusi Nirlaba dalam melakukan riset media. *Input* dari tahapan ini adalah wawancara dengan *subject matter* dan penelitian awal penulis. *Output* dari tahapan ini adalah hasil analisis sistem berjalan serta masalah-masalah yang ada.
2. *Suggestion*, yaitu menemukan solusi yang tepat agar riset media yang dilakukan subdit bersangkutan menjadi lebih efektif dan kemudian mengajukan rancangan suatu sistem aplikasi. *Input* dari tahapan ini adalah hasil analisis sistem berjalan serta masalah-masalah yang ada. *Output* dari tahapan ini adalah rancangan penulis untuk mengatasi masalah-masalah yang muncul pada sistem berjalan atau rancangan sistem usulan.
3. *Development*, yaitu mengembangkan suatu sistem aplikasi sesuai dengan rancangan pada tahapan sebelumnya. *Input* dari tahapan ini adalah rancangan sistem usulan. *Output* dari tahapan ini adalah sebuah artifak sesuai dengan rancangan sistem usulan.
4. *Evaluation*, yaitu menguji kinerja dari sistem aplikasi yang telah dibangun dengan uji coba internal, yaitu melihat apakah rancangan awal sistem usulan dibangun sudah terpenuhi, kemudian mengujicobakannya pada *subject matter* yang juga bertindak sebagai pengguna

sistem aplikasi ini serta dengan melihat perbandingan waktu yang dibutuhkan untuk melakukan riset media sebelum dan sesudah menggunakan sistem usulan dan menyajikan hasilnya dalam bentuk tabel. *Input* dari tahapan ini adalah artifak yang sudah dibangun dari tahapan sebelumnya. *Output* dari tahapan ini adalah daftar kelebihan dan kelemahan sistem.

5. *Conclusion*, yaitu menyimpulkan keefektifan dari metode penelitian yang dilakukan ini. *Input* dari tahapan ini adalah daftar kelebihan dan kelemahan sistem. *Output* dari tahapan ini adalah daftar kesimpulan dan saran.

Sumber data yang digunakan pada pembangunan sistem aplikasi ini adalah berikut.

1. *Awareness of problem*. Pada tahapan ini, yang menjadi sumber data adalah hasil wawancara dengan *subject matter* mengenai alur sistem berjalan. Selain itu, penulis melakukan pengamatan langsung di lapangan dengan menyimulasikan proses perangkuman berita dari beberapa situs. Data-data ini digunakan untuk tahapan analisis masalah yang terdapat pada sistem berjalan.
2. *Suggestion*. Pada tahapan ini, yang menjadi sumber data adalah hasil analisis pada tahap sebelumnya yang menghasilkan rincian-rincian masalah sistem berjalan.
3. *Development*. Untuk melakukan pengembangan sistem usulan, data yang diperlukan adalah pola berita-berita pada situs yang dijadikan sumber *crawling*, yaitu situs Bisnis dan Kontan. Dari pola-pola yang ada, dapat ditemukan metode untuk melakukan *crawling* berita dari situs asalnya.
4. *Evaluation*. Sumber data pada tahapan ini adalah hasil simulasi program yang menunjukkan tingkat keefektifan sistem usulan. Sumber data yang berikutnya yaitu tanggapan dari *subject matter* selaku pengguna sistem aplikasi usulan. Kedua sumber ini dapat dijadikan bahan acuan untuk mengidentifikasi kelebihan dan kelemahan sistem usulan yang telah dibuat.
5. *Conclusion*. Sumber data pada tahapan ini adalah kelebihan dan kelemahan sistem usulan seperti yang telah diidentifikasikan pada tahapan sebelumnya. Data-data tersebut dapat dijadikan bahan acuan untuk membuat kesimpulan dan saran dari sistem usulan.

IV. HASIL DAN PEMBAHASAN

Berdasarkan tahapan pembangunan *Design Oriented Research*, maka didapatkan beberapa hasil yaitu sebagai berikut.

1. *Awareness of Problem*

Pada tahapan awal penelitian ini, dilakukan analisis pada sistem berjalan yang sudah ada sebelumnya. Dari analisis yang telah dilakukan, didapatkan beberapa fakta pada sistem berjalan, yaitu sebagai berikut.

a. Keterlibatan pengguna

Pengguna harus melakukan riset media secara manual. Sebelum melakukan riset media, proses perangkuman berita yang seharusnya bisa dilakukan dengan lebih cepat malah masih dilakukan secara manual yang berdampak pada kurang efektifnya pemanfaatan waktu dan tenaga pengguna untuk melakukan pekerjaannya. Padahal seharusnya waktu yang digunakan untuk perangkuman yang dilakukan secara manual ini bisa digunakan untuk melakukan pekerjaan lain. Selain itu pengguna juga harus terus menerus memonitor situs-situs berita *online* yang diinginkan agar pekerjaannya tidak menumpuk ketika jam kerja hampir selesai. Seluruh proses ini cukup memakan waktu kerja.

b. Proses perangkuman

Berdasarkan hasil pengamatan penulis, rata-rata setiap 45 menit sekali ada satu berita ekonomi baru yang muncul pada satu situs berita ekonomi *online*. Artinya dalam satu hari ada sekitar kurang lebih 30 berita ekonomi pada satu situs berita ekonomi *online*. Untuk dua situs berita ekonomi *online*, maka total berita yang ada bisa mencapai kira-kira 60 berita ekonomi setiap harinya. Sementara itu, waktu yang dibutuhkan untuk menyalin (*copy-paste*) sebuah berita ke dalam *format Microsoft Word* rata-rata sekitar 1-2 menit per berita, jika ditambah dengan waktu untuk membaca sekilas (*skimming*) terlebih dahulu berita-berita tersebut (untuk meminimalisir berita yang mirip dari situs yang berbeda disalin dua kali) maka kira-kira untuk satu berita dibutuhkan waktu 3-4 menit. Kemudian untuk menyeragamkan *format* berita yang telah dikumpulkan agar menjadi lebih rapi untuk dibaca (asumsinya penyeragaman dilakukan setelah seluruh berita dalam satu hari tersebut dikumpulkan) maka membutuhkan waktu kurang lebih 10 menit. Berarti dalam setiap harinya, total waktu yang dibutuhkan untuk mengerjakan seluruh rangkuman berita tersebut adalah sekitar (3*60) + 10 = 190 menit = 3,2 jam. Sedangkan *subject matter* yang juga adalah pengguna sistem aplikasi ini cukup kesulitan dalam mengalokasikan waktu untuk melakukan setiap pekerjaannya.

c. Pembacaan berita *streaming*

Jika pengguna ingin membaca berita dari situs-situs berita ekonomi *online* yang berbeda, maka pengguna harus membuka situs-situs tersebut satu persatu dan membaca berita pada situs-situs tersebut. Hal ini menyebabkan pengguna harus berpindah dari *tab* atau *window browser* yang satu ke yang lain dan hal ini tentunya akan memakan waktu yang cukup lama jika dilakukan secara berulang-ulang.

Selain itu, kebutuhan-kebutuhan pengguna adalah sebagai berikut:

1. Mencari dan menemukan suatu metode yang dapat melakukan perangkuman berita dari situs-situs berita ekonomi *online* secara otomatis dalam format *Microsoft Word* setiap hari untuk berbagai kepentingan, antara lain untuk keperluan pembuatan riset media, agar dapat mengetahui fenomena ekonomi secara cepat, serta membantu orang-orang di subdirektorat untuk dapat membaca bermacam-macam berita yang berkaitan dengan ekonomi dalam waktu singkat.
2. Mencari dan menemukan suatu metode yang dapat meminimalisasi kemiripan dari berita-berita yang dirangkum, sehingga jika ada dua buah atau lebih berita yang sama, yang akan dirangkum hanya salah satu berita dan berita-berita yang mirip lainnya tidak akan dimasukkan ke dalam perangkuman.
3. Membuat rangkuman berita dengan cara mengambil kalimat yang hanya berisi angka saja untuk masing-masing berita asli.

2. *Suggestion*

Dari analisis masalah di atas, maka penulis menemukan sebuah solusi untuk memudahkan pekerjaan pengguna dalam melakukan perangkuman berita dari situs-situs berita ekonomi *online* dengan mengajukan rancangan suatu sistem aplikasi yang secara garis besar memiliki fungsi-fungsi berikut.

1. Meminimalisasi ketelibatan pengguna dengan sistem dengan cara melakukan sebagian besar pekerjaan secara otomatis.
2. Melakukan perangkuman berita-berita ekonomi yang berada dalam situs-situs berita ekonomi *online* yang sudah ditentukan sebelumnya oleh pengguna dalam *format Microsoft Word* setiap hari, yang sebelumnya berita-berita yang mirip dari dua atau lebih situs yang berbeda hanya akan diambil salah satu sehingga meminimalisasi pengulangan. Berikut merupakan contoh tahapan-tahapan pengecekan kesamaan dari dua dokumen dengan menggunakan metode *cosine similarity*.

- Misalkan dokumen A (d_A) dan dokumen B (d_B) berisi kalimat sebagai berikut.

$$d_A = \text{"}Christopher\ has\ two\ green\ apples\text{"} \quad (1)$$

$$d_B = \text{"}Christopher\ has\ two\ red\ apples\text{"} \quad (2)$$

- Bentuk vektor dari masing-masing dokumen tersebut.

$$\vec{V}(d_A) = [\text{"}Christopher\text{"}, \text{"}has\text{"}, \text{"}two\text{"}, \text{"}green\text{"}, \text{"}apples\text{"}] \quad (3)$$

$$\vec{V}(d_B) = [\text{"}Christopher\text{"}, \text{"}has\text{"}, \text{"}two\text{"}, \text{"}red\text{"}, \text{"}apples\text{"}] \quad (4)$$

- Bentuk vektor gabungan dari vektor (4) dan vektor (5) di atas.

$$\vec{V}(d_A d_B) = ["Christopher", "has", "two", "green", "apples", "red"] \quad (5)$$

- Bentuk vektor frekuensi kemunculan kata-kata berdasarkan vektor gabungan (6) untuk masing-masing vektor.

$$\vec{V}_F(d_A) = [1,1,1,1,1,0] \quad (6)$$

$$\vec{V}_F(d_A) = [1,1,1,0,1,1] \quad (7)$$

- Lakukan perkalian titik (dot product) untuk kedua vektor frekuensi (6) dan (7) di atas.

$$\vec{V}_F(d_A) \cdot \vec{V}_F(d_B) = [1,1,1,1,1,0] \cdot [1,1,1,0,1,1]$$
$$= [(1 \cdot 1) + (1 \cdot 1) + (1 \cdot 1) + (1 \cdot 0) + (1 \cdot 1) + \quad (0 \cdot 1)]$$
$$= [1 + 1 + 1 + 0 + 1 + 0]$$
$$= 4 \quad (8)$$

- Hitung panjang *Euclidean* dari masing-masing vektor dokumen.

$$\left|\vec{V}(d_A)\right| = \sqrt{1^2 + 1^2 + 1^2 + 1^2 + 1^2 + 0^2} = \sqrt{5} \quad (9)$$

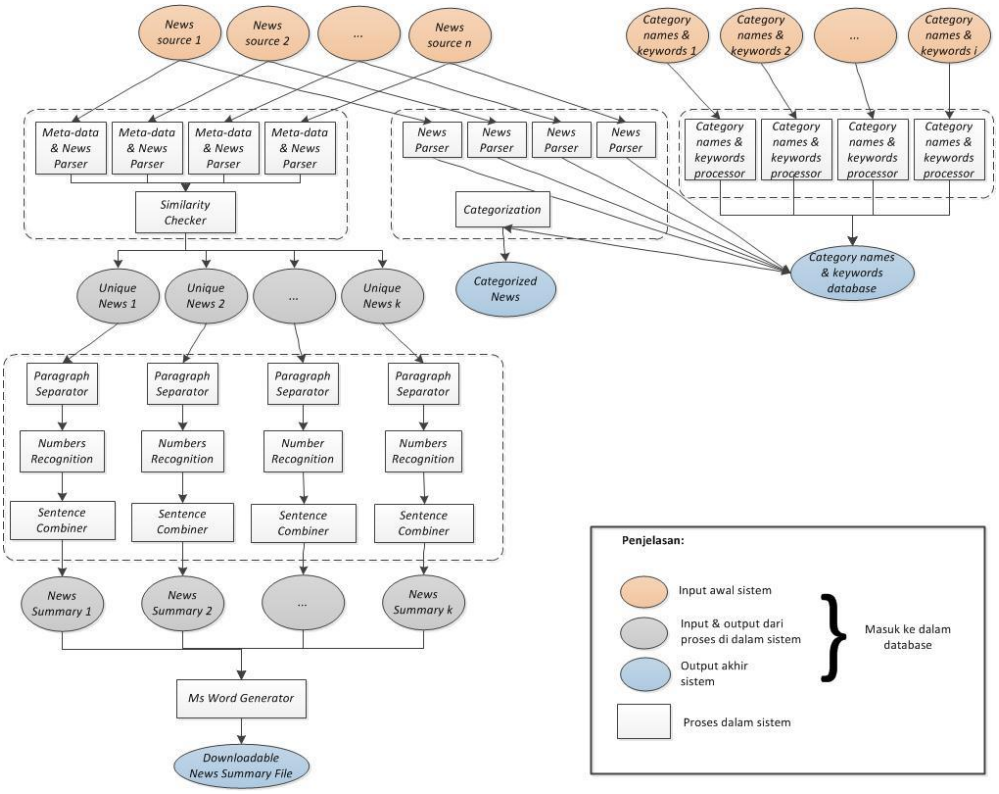
$$\left|\vec{V}(d_B)\right| = \sqrt{1^2 + 1^2 + 1^2 + 0^2 + 1^2 + 1^2} = \sqrt{5} \quad (10)$$

- Hitung nilai *cosine similarity* kedua dokumen tersebut
- $sim(d_1, d_2) = \frac{\vec{V}(d_1) \cdot \vec{V}(d_2)}{\left|\vec{V}(d_1)\right|\left|\vec{V}(d_2)\right|'} = \frac{4}{\sqrt{5} \times \sqrt{5}} = 0{,}8$

Maka nilai *cosine similarity* dari dokumen A dan dokumen B adalah 0,8.

3. Memungkinkan pengguna melihat berita dari berbagai situs-situs berita ekonomi *online* yang sudah ditentukan sebelumnya oleh pengguna secara terintegrasi pada satu tempat dan bersifat *real-time* dan menyediakan fasilitas penambahan kategori berdasarkan kata kunci sehingga berita-berita yang terintegrasi tersebut dapat dicari berdasarkan kategori, waktu, serta sumber berita.

Arsitektur Sistem



Gambar 2. Arsitektur sistem

Penjelasan dari arsitektur sistem di atas dapat dilihat pada tabel 1 berikut ini:

Tabel 1. Penjelasan arsitektur sistem dan teknologi yang digunakan

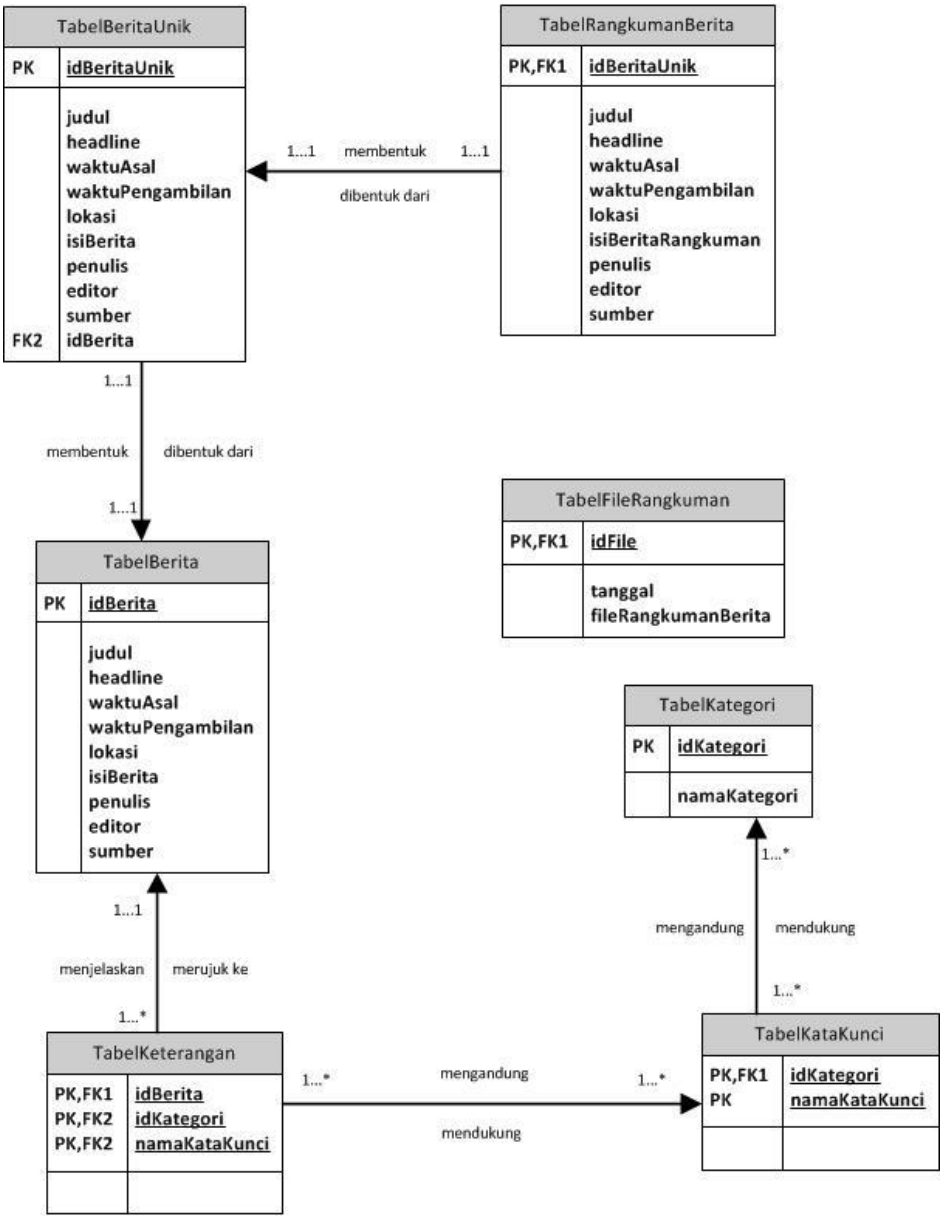
Nama Item	Penjelasan	Teknologi yang Digunakan
News source 1, 2, ..., n	Item ini merupakan representasi dari berita-berita yang tersebar di situs-situs berita ekonomi *online* yang telah ditentukan atau disebut juga berita mentah. Ketika sistem mengambil berita ini secara otomatis, maka isi berita-berita ini akan masuk ke dalam basis data. Item ini merupakan *input* untuk proses *Meta-data & News Parser* dan *News Parser*	*Crawler*
Meta-data & News Parser	Item ini merupakan proses yang mengolah *News Source*. Pada proses ini terjadi pemisahan-pemisahan elemen-elemen berita, yaitu judul, *headline*, waktu asal berita, isi berita, penulis, editor, sumber berita, serta waktu pengambilan berita dari situs oleh sistem	*Wrapper, jsoup library*

Nama Item	Penjelasan	Teknologi yang Digunakan
Similarity Checker	Proses ini merupakan proses pengecekan kesamaan antara dua berita. Berita-berita yang ada saling dibandingkan satu dengan yang lain. *Output* dari proses ini adalah *Unique News*	*Cosine Similarity*
Unique News 1, 2, ..., k	Item ini menerima *input* dari proses sebelumnya yaitu *Similarity Checker*. *Unique News* merupakan berita-berita yang unik yang sudah tidak memiliki kesamaan satu dengan yang lainnya. Item ini merupakan *input* untuk proses selanjutnya yaitu *Paragraph Separator*. Nilai k adalah kurang dari atau sama dengan n	Tidak ada
Paragraph Separator	Proses ini melakukan pemisahan isi berita *Unique News* dari yang semula merupakan satu kesatuan paragraf menjadi kalimat-kalimat individu yang dimasukkan dalam sebuah *array*	Tidak ada
Numbers Recognition	Proses ini membaca kalimat-kalimat yang telah terbentuk dari proses sebelumnya. Pada proses ini, kalimat-kalimat yang mengandung angka tetap disimpan dan kalimat-kalimat yang tidak mengandung angka dibuang atau tidak akan dipergunakan untuk proses selanjutnya	Tidak ada
Sentence Combiner	Proses ini merupakan kelanjutan dari proses sebelumnya. Pada proses ini terjadi penggabungan kalimat-kalimat yang mengandung angka (seperti yang terjadi pada proses sebelumnya) menjadi satu kesatuan paragraf yang baru. *Output* dari proses ini adalah *News Summary*	Tidak ada
News Summary 1, 2, ..., k	Item ini merupakan *output* dari rangkaian proses sebelumnya. Item ini merupakan rangkuman dari *input* awal sistem yaitu *News Source* (berita mentah dari internet).	Tidak ada
Ms. Word Generator	Pada proses ini terjadi konversi berita dari basis data ke dalam *format Micosoft Word*. Pada proses ini juga terjadi perapihan dan penyamaan *format* berita sehingga gampang dan nyaman dibaca oleh *user*	*Apache POI library*
Downloadable News Summary File	Item ini merupakan *output* akhir dari sistem, yang merupakan *file* rangkuman berita yang secara otomatis masuk ke dalam komputer pengguna atau bisa diunduh sendiri. *File* ini memiliki *format Microsoft Word*	Tidak ada

Nama Item	Penjelasan	Teknologi yang Digunakan
News Parser	Proses ini berbeda dengan proses *Meta-data & News Parser*. Pada proses ini terjadi pembacaan dan pencocokan kata-kata yang terdapat dalam berita dengan kata kunci yang ada dalam basis data	Tidak ada
Categorization	Proses ini merupakan kelanjutan dari proses sebelumnya, dimana proses ini mengkategorikan berita-berita yang ada sesuai dengan proses pencocokan berita yang terjadi sebelumnya. *Output* dari proses ini adalah *Categorized News*	Tidak ada
Categorized News	Item ini merupakan *output* dari proses sebelumnya. Item ini merupakan berita yang sudah terkategori dan ditampilkan secara *real-time* pada sistem aplikasi	Tidak ada
Category Names & Keywords 1, 2, ..., i	Item ini merupakan *input* untuk sistem aplikasi ini. *Input* ini diberikan oleh pengguna	Tidak ada
Category Names and Keywords Processor	Pada proses ini terjadi validasi kata kunci-kata kunci dan kategori yang dimasukkan oleh pengguna sehingga dapat diproses lebih lanjut oleh sistem	Tidak ada
Category Names & Keywords Database	Item ini merupakan basis data yang menyimpan kata kunci-kata kunci dan kategori yang dimasukkan oleh pengguna yang kemudian akan dibandingkan oleh berita-berita yang telah diambil dari internet	Tidak ada

Rancangan Basis Data

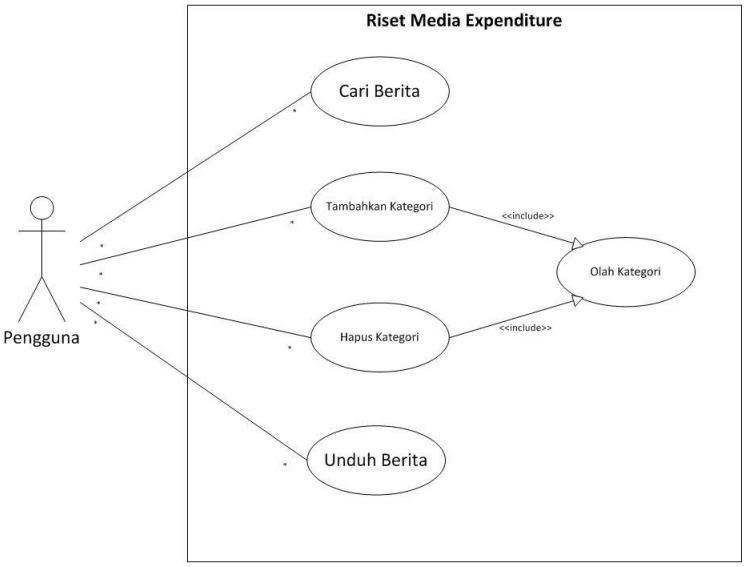
Gambar 3 merupakan ERD (*Entity Relationship Diagram*) dari rancangan logik basis data, yang menunjukkan entitas yang digunakan dalam sistem yang akan dibangun.



Gambar 3. ***Entity relationship diagram*** **rancangan logik basis data**

Use Case Diagram

Gambar 4 merupakan diagram *use case*, yang menggambarkan *actor* (pelaku/pengguna) dengan masing-masing kasus penggunaan dalam sistem yang akan dibangun.



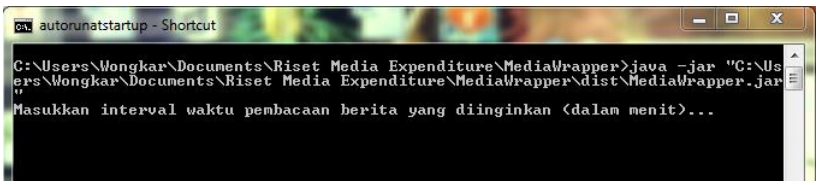
Gambar 4. ***Use case diagram***

3. *Development*

Dari rancangan dan pemodelan tersebut, penulis telah membangun sebuah sistem aplikasi yang dapat mengakomodir kegiatan riset media seperti yang telah diuraikan di atas.

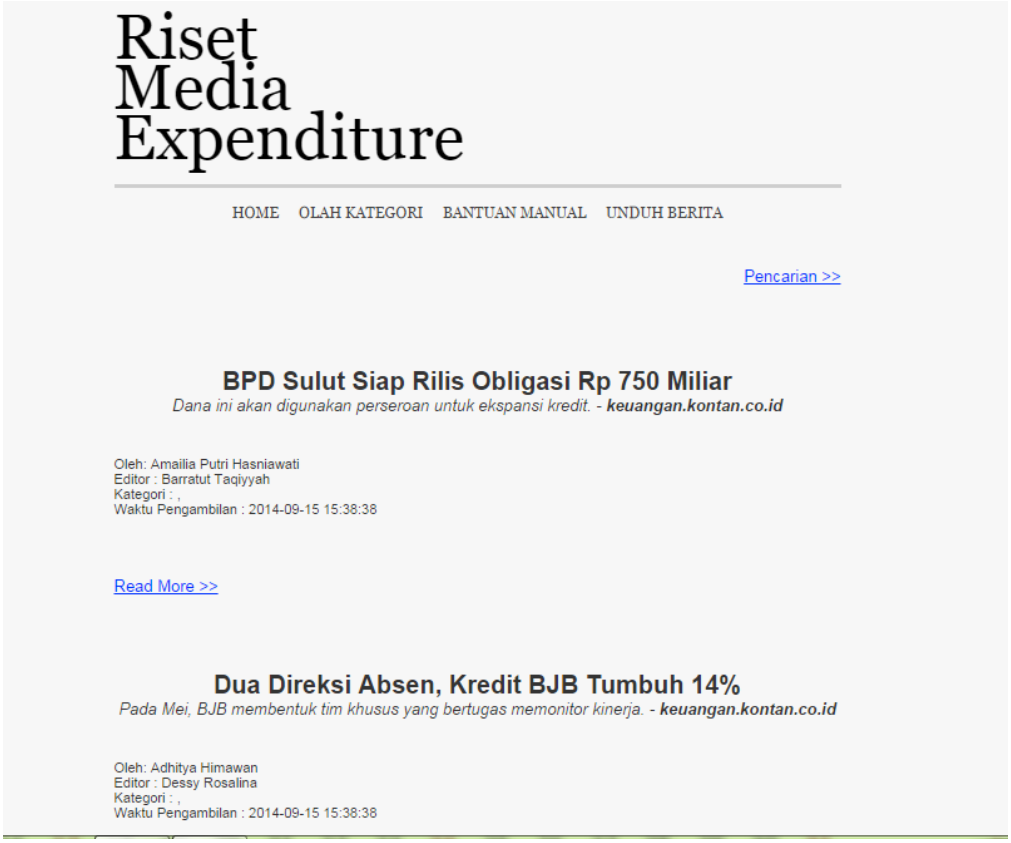
Sistem aplikasi ini terdiri atas dua bagian, yaitu bagian *back-end* dan bagian *front-end*. Bagian *back-end* yang berjalan di belakang layar ini dikembangkan dengan bahasa pemrograman *Java*. Bagian ini tidak memiliki *graphical user interface* (GUI) sehingga proses yang terjadi pada bagian ini hanya dapat dilihat atau dipantau dengan output baris-baris yang berupa *command line*. Bagian ini berfungsi untuk mengambil berita-berita dari situs asalnya, mengecek kesamaan, memilih kalimat yang hanya mengandung angka, serta membuat *file* rangkuman dalam format *Microsoft Word*. Bagian ini dijalankan pada saat *startup* komputer, sehingga ketika komputer pengguna dinyalakan, maka program akan langsung mulai melakukan *crawling* dan *wrapping* berita. Proses *crawling* dan *wrapping* ini berlangsung secara berulang-ulang pada interval tertentu yang ditentukan oleh pengguna dan berhenti pada pukul 15.00 ketika jam kerja akan berakhir. Pada pukul 15.00, bagian ini akan melakukan pengecekan kesamaan, pemilihan kalimat yang mengandung angka, dan pembentukan *file* rangkuman dalam format Microsoft Word sehingga *file* yang terbentuk merupakan hasil dari *crawling* dan *wrapping* berita-berita sejak komputer dijalankan sampai pukul 15.00.

Gambar 5 di bawah ini merupakan *screenshot* tampilan *command line* dari bagian *back-end*.



Gambar 5. *Screenshot* sistem usulan bagian *back-end*

Bagian *front-end* dikembangkan dengan bahasa pemrograman PHP dan *Javascript* serta menggunakan *framework* Yii sehingga pengguna dapat melihat *output* yang dihasilkan oleh bagian *back-end* dalam bentuk halaman *web*. Berita-berita asli (yang bukan berupa rangkuman) yang telah diambil dari situs aslinya dapat dilihat pada halaman *web* tersebut. Pada bagian *front-end* ini juga pengguna dapat melakukan olah kategori (hapus dan tambahkan kategori). Salah satu *screenshot* dari bagian *front-end* dapat dilihat pada Gambar 6 berikut.



Gambar 6. *Screenshot* sistem usulan menu *home*

4. *Evaluation*

Pada tahapan ini dilakukan evaluasi terhadap sistem yang sudah dibangun. Rincian evaluasi yang ada disesuaikan dengan rincian *suggestion* yang sudah dibuat sebelumnya.

1. Keterlibatan pengguna

Pada poin yang pertama ini, dapat dilihat perbedaan keterlibatan pengguna sebelum dan sesudah menggunakan sistem aplikasi yang dibangun. Perbedaan ini dapat dilihat pada tabel 2 di bawah ini.

Tabel 2. Perbandingan total aktivitas dan total waktu antara sistem berjalan dengan sistem usulan (dengan jumlah pengguna 1 orang)

<table>
<tr><td colspan="6">Aktivitas yang dilakukan pengguna dan sistem</td></tr>
<tr><td>Sistem Berjalan</td><td>Estimasi Waktu yang Dibutuhkan</td><td>Total Waktu</td><td>Sistem Usulan</td><td>Estimasi Waktu yang Dibutuhkan</td><td>Total Waktu</td></tr>
<tr><td>Membuka situs berita Bisnis</td><td>30 detik</td><td>30 detik</td><td rowspan="2">*Crawling* situs berita Bisnis dan Kontan dan *wrap* berita yang baru ada setiap 15 menit sekali *)</td><td rowspan="2">@1 detik</td><td rowspan="2">untuk 7 jam dalam sehari (jam 08.00-15.00), 28 kali = 28 detik</td></tr>
<tr><td>Membuka situs berita Kontan</td><td>30 detik</td><td>30 detik</td></tr>
<tr><td>Membuka *link* berita 1, 2, ..., n pada situs berita Bisnis</td><td>@60 detik</td><td>60 detik x 30 berita (kurang lebih) = 1800 detik</td><td rowspan="2">Membentuk *file* XML dari berita-berita yang telah di-*wrap* *)</td><td rowspan="2">@1 detik</td><td rowspan="2">1 detik x 60 berita (kurang lebih) = 60 detik</td></tr>
<tr><td>Membuka *link* berita 1, 2, ..., n pada situs berita Kontan</td><td>@60 detik</td><td>60 detik x 30 berita (kurang lebih) = 1800 detik</td></tr>
<tr><td>*Skimming* berita 1, 2, ..., n pada situs berita Bisnis</td><td>@120 detik</td><td>120 detik x 30 berita (kurang lebih) = 3600 detik</td><td rowspan="2">Memeriksa kesamaan dari berita-berita yang telah ada *)</td><td rowspan="2"></td><td rowspan="2">Untuk kombinasi 60 berita sekitar 360 detik</td></tr>
<tr><td>*Skimming* berita 1, 2, ..., n pada situs berita Kontan</td><td>@120 detik</td><td>120 detik x 30 berita (kurang lebih) = 3600 detik</td></tr>
</table>

Aktivitas yang dilakukan pengguna dan sistem					
Sistem Berjalan	Estimasi Waktu yang Dibutuhkan	Total Waktu	Sistem Usulan	Estimasi Waktu yang Dibutuhkan	Total Waktu
Menyalin-tempel (*copy-paste*) berita 1, 2, ..., n dari situs berita Bisnis ke dalam *file Microsoft Word*	@60 detik	60 detik x 30 berita (kurang lebih) = 1800 detik	Membuat rangkuman berita *)	@2 detik	1 detik x 60 berita (kurang lebih) = 60 detik
Menyalin-tempel (*copy-paste*) berita 1, 2, ..., n dari situs berita Kontan ke dalam *file Microsoft Word*	@60 detik	60 detik x 30 berita (kurang lebih) = 1800 detik			
Merapikan dan menyamakan *format-format* dalam *file Microsoft Word*	600 detik	600 detik	Membentuk *file Microsoft Word* dari rangkuman-rangkuman berita tersebut *)	2 detik	2 detik
Total Aktivitas: 9		**Total Waktu: 15.060 detik ≈ 251 menit ≈ 4,2 jam**	**Total Aktivitas: 5**		**Total Waktu: 510 detik ≈ 8,5 menit**

Keterangan: *) dilakukan oleh sistem aplikasi

Dari tabel 2 di atas dapat dilihat bahwa perbandingan total aktivitas dan total waktu pada sistem berjalan dan sistem usulan cukup berbeda secara signifikan. Total aktivitas yang dilakukan oleh sistem usulan hampir setengah dari total aktivitas yang dilakukan oleh sistem berjalan, yaitu 5 : 9 (55,56%, terjadi reduksi sebesar 44,4%). Artinya, sistem usulan mampu mereduksi aktivitas yang dilakukan oleh sistem berjalan sebesar hampir 50%. Namun jika kita telaah lebih lanjut lagi, seluruh aktivitas yang terjadi pada sistem usulan merupakan aktivitas

yang dilakukan secara otomatis oleh sistem aplikasi. Sehingga dapat disimpulkan bahwa sistem usulan mereduksi keterlibatan pengguna hingga hampir 100%.

2. Proses perangkuman

Jika kita mengamati tabel 2 yang sebelumnya, dapat dilihat bahwa dari sisi waktu, perbandingan antara total waktu yang digunakan oleh sistem usulan dengan sistem berjalan adalah 8,5 : 251 atau sekitar 3,386%. Dengan demikian, dapat disimpulkan bahwa sistem usulan mereduksi total waktu yang diperlukan untuk membuat rangkuman berita hingga 96,614%.

Sebagai perbandingan dari segi estimasi total waktu yang dibutuhkan dengan sumber daya manusia yang diberdayakan, maka disajikan juga tabel yang menggambarkan total waktu yang dilakukan oleh pengguna pada sistem berjalan berdasarkan jumlah sumber daya manusia yang terlibat.

Tabel 3. Perbandingan total aktivitas dan total waktu antara sistem berjalan dengan sistem usulan (dengan jumlah pengguna 2 orang)

<table>
<tr><td colspan="6">Aktivitas yang dilakukan pengguna dan sistem (2 orang)</td></tr>
<tr><td>Sistem Berjalan</td><td>Estimasi Waktu yang Dibutuhkan</td><td>Total Waktu</td><td>Sistem Usulan</td><td>Estimasi Waktu yang Dibutuhkan</td><td>Total Waktu</td></tr>
<tr><td>Membuka situs berita Bisnis</td><td>30 detik</td><td>30 detik</td><td rowspan="2">Crawling situs berita Bisnis dan Kontan dan wrap berita yang baru ada setiap 15 menit sekali *)</td><td rowspan="2">@1 detik</td><td rowspan="2">untuk 7 jam dalam sehari (jam 08.00-15.00), 28 kali = 28 detik</td></tr>
<tr><td>Membuka situs berita Kontan</td><td>30 detik</td><td>30 detik</td></tr>
<tr><td>Membuka link berita 1, 2, ..., n pada situs berita Bisnis</td><td>@60 detik</td><td>60 detik x 15 berita / orang (kurang lebih) = 900 detik</td><td rowspan="2">Membentuk file XML dari berita-berita yang telah di-wrap *)</td><td rowspan="2">@1 detik</td><td rowspan="2">1 detik x 60 berita (kurang lebih) = 60 detik</td></tr>
<tr><td>Membuka link berita 1, 2, ..., n pada situs berita Kontan</td><td>@60 detik</td><td>60 detik x 15 berita / orang (kurang lebih) = 900 detik</td></tr>
</table>

Aktivitas yang dilakukan pengguna dan sistem (2 orang)					
Sistem Berjalan	Estimasi Waktu yang Dibutuhkan	Total Waktu	Sistem Usulan	Estimasi Waktu yang Dibutuhkan	Total Waktu
Skimming berita 1, 2, ..., n pada situs berita Bisnis	@120 detik	120 detik x 15 berita / orang (kurang lebih) = 1800 detik	Memeriksa kesamaan dari berita-berita yang telah ada *)		Untuk kombinasi 60 berita sekitar 360 detik
Skimming berita 1, 2, ..., n pada situs berita Kontan	@120 detik	120 detik x 15 berita / orang (kurang lebih) = 1800 detik			
Menyalin-tempel (*copy-paste*) berita 1, 2, ..., n dari situs berita Bisnis ke dalam *file Microsoft Word*	@60 detik	60 detik x 15 berita / orang (kurang lebih) = 900 detik	Membuat rangkuman berita *)	@2 detik	1 detik x 60 berita (kurang lebih) = 60 detik
Menyalin-tempel (*copy-paste*) berita 1, 2, ..., n dari situs berita Kontan ke dalam *file Microsoft Word*	@60 detik	60 detik x 15 berita / orang (kurang lebih) = 900 detik			
Merapikan dan menyamakan *format-format* dalam *file Microsoft Word*	600 detik	600 detik	Membentuk *file Microsoft Word* dari rangkuman-rangkuman berita tersebut *)	2 detik	2 detik
Total Aktivitas: 9		**Total Waktu: 7.860 detik ≈ 131 menit ≈ 2,18 jam**	**Total Aktivitas: 5**		**Total Waktu: 510 detik ≈ 8,5 menit**

Keterangan: *) dilakukan oleh sistem aplikasi

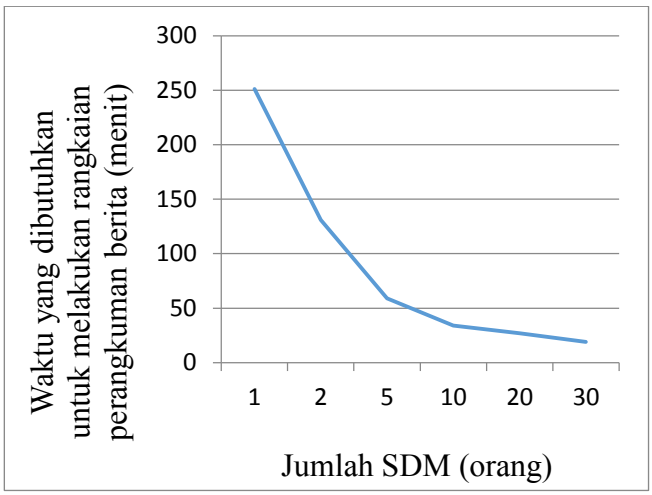
Tabel 4. Perbandingan total aktivitas dan total waktu antara sistem berjalan dengan sistem usulan (dengan jumlah pengguna 5 orang)

<table>
<tr><td colspan="6">Aktivitas yang dilakukan pengguna dan sistem (5 orang)</td></tr>
<tr><td>Sistem Berjalan</td><td>Estimasi Waktu yang Dibutuhkan</td><td>Total Waktu</td><td>Sistem Usulan</td><td>Estimasi Waktu yang Dibutuhkan</td><td>Total Waktu</td></tr>
<tr><td>Membuka situs berita Bisnis</td><td>30 detik</td><td>30 detik</td><td rowspan="2">*Crawling* situs berita Bisnis dan Kontan dan *wrap* berita yang baru ada setiap 15 menit sekali *)</td><td rowspan="2">@1 detik</td><td rowspan="2">untuk 7 jam dalam sehari (jam 08.00-15.00), 28 kali = 28 detik</td></tr>
<tr><td>Membuka situs berita Kontan</td><td>30 detik</td><td>30 detik</td></tr>
<tr><td>Membuka *link* berita 1, 2, ..., n pada situs berita Bisnis</td><td>@60 detik</td><td>60 detik x 6 berita / orang (kurang lebih) = 360 detik</td><td rowspan="2">Membentuk *file* XML dari berita-berita yang telah di-*wrap* *)</td><td rowspan="2">@1 detik</td><td rowspan="2">1 detik x 60 berita (kurang lebih) = 60 detik</td></tr>
<tr><td>Membuka *link* berita 1, 2, ..., n pada situs berita Kontan</td><td>@60 detik</td><td>60 detik x 6 berita / orang (kurang lebih) = 360 detik</td></tr>
<tr><td>*Skimming* berita 1, 2, ..., n pada situs berita Bisnis</td><td>@120 detik</td><td>120 detik x 6 berita / orang (kurang lebih) = 720 detik</td><td rowspan="2">Memeriksa kesamaan dari berita-berita yang telah ada *)</td><td rowspan="2"></td><td rowspan="2">Untuk kombinasi 60 berita sekitar 360 detik</td></tr>
<tr><td>*Skimming* berita 1, 2, ..., n pada situs berita Kontan</td><td>@120 detik</td><td>120 detik x 6 berita / orang (kurang lebih) = 720 detik</td></tr>
<tr><td>Menyalin-tempel (*copy-paste*) berita 1, 2, ..., n dari situs berita Bisnis ke dalam *file Microsoft Word*</td><td>@60 detik</td><td>60 detik x 6 berita / orang (kurang lebih) = 360 detik</td><td>Membuat rangkuman berita *)</td><td>@2 detik</td><td>1 detik x 60 berita (kurang lebih) = 60 detik</td></tr>
</table>

Aktivitas yang dilakukan pengguna dan sistem (5 orang)					
Sistem Berjalan	Estimasi Waktu yang Dibutuhkan	Total Waktu	Sistem Usulan	Estimasi Waktu yang Dibutuhkan	Total Waktu
Menyalin-tempel (*copy-paste*) berita 1, 2, ..., n dari situs berita Kontan ke dalam *file Microsoft Word*	@60 detik	60 detik x 6 berita / orang (kurang lebih) = 360 detik			
Merapikan dan menyamakan *format-format* dalam *file Microsoft Word*	600 detik	600 detik	Membentuk *file Microsoft Word* dari rangkuman-rangkuman berita tersebut *)	2 detik	2 detik
Total Aktivitas: 9		**Total Waktu: 3.540 detik ≈ 59 menit ≈ 0,98 jam**	**Total Aktivitas: 5**		**Total Waktu: 510 detik ≈ 8,5 menit**

Keterangan: *) dilakukan oleh sistem aplikasi

Dari gambar 7 dapat dilihat bahwa penurunan total waktu untuk melakukan rangkaian proses perangkuman berita terjadi secara signifikan jika jumlah sumber daya manusia yang melakukan proses perangkuman berita tersebut berubah dari 1 orang menjadi 2 orang. Untuk jumlah sumber daya manusia yang lainnya yaitu 5 orang, 10 orang, 20 orang, dan 30 orang, perubahan total waktu yang terjadi kurang signifikan. Hal ini terjadi karena ada beberapa proses perangkuman berita yang tidak bisa dibagi sesuai jumlah orang, seperti membuka situs berita dan merapikan dan menyamakan format dalam *file Microsoft Word*. Untuk diketahui bahwa dalam perhitungan perbandingan sumber daya manusia ini, jumlah berita dalam satu hari dari dua situs berita ekonomi *online* yang berbeda dibagi sesuai jumlah orang yang mengerjakannya.



Gambar 7. Perbandingan waktu antara sistem usulan dengan sistem berjalan untuk melakukan rangkaian proses perangkuman berita dengan jumlah sumber daya manusia (SDM) yang bervariasi

Selain itu, dari gambar 7 dapat juga dilihat bahwa untuk jumlah sumber daya manusia sebanyak 30 orang, yang berarti bahwa 1 orang hanya perlu mengerjakan satu berita saja dari masing-masing situs berita ekonomi online, estimasi total waktu yang dibutuhkan untuk melakukan rangkaian proses perangkuman tersebut adalah 19 menit. Jika dibandingkan dengan estimasi total waktu yang dilakukan oleh sistem usulan, yaitu sebesar 8,5 menit, tentunya angka tersebut masih cukup besar. Dengan demikian, dapat disimpulkan bahwa dengan jumlah sumber daya manusia sebanyak 30 orang masih belum dapat mengimbangi efisiensi sistem usulan dari segi waktu.

3. Pembacaan berita *streaming*

   Berita-berita dari situs berita ekonomi *online* akan dibentuk dalam *format* XML, kemudian *file* XML tersebut akan dipindahkan ke dalam basis data.

5. *Conclusion*

Dari evaluasi yang telah dilakukan, kelebihan yang dimiliki oleh sistem aplikasi ini adalah sebagai berikut:

a. Sistem sudah bisa melakukan perangkuman berita secara otomatis dari situs-situs berita ekonomi *online* yang ditentukan oleh pengguna, dengan meminimalisir berita yang mirip satu dengan yang lain.
b. Sistem sudah bisa mengintegrasikan berita-berita *online* dari situs-situs yang berbeda ke dalam sistem aplikasi yang telah dibuat sehingga dapat dibaca pada satu tempat.

c. Sistem sudah bisa membantu pengguna untuk mencari berita yang telah terambil berdasarkan kategori, waktu, dan sumber berita.

Namun, masih ada beberapa kelemahan yang dimiliki oleh sistem aplikasi ini, yaitu:

a. Metode perangkuman berita yang hanya mengandung angka belum dilakukan dengan memperhatikan arti semantik.
b. Situs-situs berita ekonomi *online* yang diambil belum terlalu banyak.
c. *Wrapper* yang dibuat masih secara manual dan belum dinamis sehingga ketika *layout* dari halaman *web* yang ditentukan sebagai sumber *wrapping* berubah, elemen-elemen di dalamnya akan gagal diambil. Selain itu, *wrapper* yang dibuat secara manual ini menyebabkan kesulitan dalam mengakomodir jumlah situs yang banyak.

Secara keseluruhan, sistem aplikasi ini sudah dapat membantu pengguna untuk menyelesaikan pekerjaannya dengan melakukan perangkuman berita yang diawali dari *crawling* situs berita ekonomi *online*, *wrapping* berita, dan meminimalisasi kesamaan dari dua atau lebih berita.

V. KESIMPULAN DAN SARAN

Kesimpulan

Kesimpulan yang didapatkan dari penelitian ini adalah sebagai berikut.

a. Sistem aplikasi sudah mampu mengurangi waktu dan tenaga yang dibutuhkan untuk melakukan perangkuman berita. Waktu yang digunakan untuk perangkuman berita direduksi sebesar kurang lebih 96,614%. Sementara itu, tenaga yang dinyatakan dengan total aktivitas yang dilakukan pengguna untuk melakukan perangkuman mulai dari membuka halaman situs berita ekonomi *online* sampai membentuk rangkuman dalam format *Microsoft Word* direduksi hingga 44,4%.
b. Sistem aplikasi sudah bisa membantu pengguna untuk membuat rangkuman berita dalam format *Microsoft Word* setiap hari.
c. Sistem aplikasi sudah bisa membantu pengguna untuk membaca berita-berita online secara terintegrasi pada satu tempat.
d. Sistem aplikasi sudah bisa membantu pengguna untuk menemukan berita yang telah terambil yang menjadi kehendaknya berdasarkan kategori, waktu, dan sumber berita.
e. Sistem aplikasi sudah bisa membantu pengguna untuk pembuatan riset media serta pembacaan berita ekonomi secara cepat dan dalam waktu singkat dengan melakukan perangkuman berita secara otomatis mulai dari *crawling* situs berita ekonomi *online*, *wrapping* berita, meminimalisasi kesamaan dari dua atau lebih berita, sampai pada perangkuman berita.

Saran

Adapun saran-saran yang dapat diberikan untuk penelitian selanjutnya adalah sebagai berikut.

a. Dapat dilakukan penelitian lebih lanjut mengenai metode penggabungan kalimat yang berisi angka saja, sehingga rangkaian kalimat yang terbentuk memiliki arti yang baku sebagai satu kesatuan paragraf.
b. Dapat diimplementasikan *wrapper* yang bersifat dinamis (*unsupervised wrapper*) sehingga bisa digunakan pada situs-situs yang berbeda.
c. Daftar situs-situs berita *online* yang ada dapat lebih dikembangkan, bukan hanya dalam bidang ekonomi saja namun dalam bidang ilmu lainnya. Selain itu kuantitasnya juga dapat ditambah menjadi lebih dari dua situs sehingga cakupan beritanya makin luas.

DAFTAR PUSTAKA

Badiyanto. (2013). *Buku Pintar Framework Yii: Cara Mudah Membangun Aplikasi Web PHP*. Jakarta: Gramedia.

Departemen Pendidikan Nasional. (2002). *Kamus Besar Bahasa Indonesia (Edisi Ketiga)*. Jakarta: Balai Pustaka.

Huang, A. (2008). Similarity Measures for Text Document Clustering. *Proceedings of the New Zealand Computer Science Research Student Conference, Christchurch, New Zealand*. 9 April 2014.http://www.milanmirkovic.com/wp-content/uploads/2012/10/pg049_Similarity_Measures_for_Text_Document_Clustering.pdf

Liu, B. (2011). *Web Data Mining: Exploring Hyperlinks, Content, and Usage Data (2nd edition)*. New York: Springer-Verlag Berlin Heidelberg.

Jackson, J. C. (2007). *Web Technologies: A Computer Science Perspective*. New Jersey: Pearson Education, Inc.

Manning, C. D., Raghavan, P., Schutze, H. (2008). *Introduction to Information Retrieval (2nd edition)*. New York: Cambridge University Press.

Nurudin (2007), Pengantar Komunikasi Massa, Raja Grafindo Persada, Jakarta.

Rizkiansyah, N. (2010). *Pengembangan Aplikasi Pendeteksi Plagiarisme Tulisan Ilmiah pada Sistem Informasi Terpadu STIS* [Skripsi]. Jakarta: Sekolah Tinggi Ilmu Statistik.

Safitri, K. (12 Oktober 2010). Pengaruh Waspada Online terhadap Pengetahuan Politik Pembacanya (Studi Korelasi pada Komunitas Waspada Online). *USU Institutional Repository*, 6. 4 September 2014.

http://repository.usu.ac.id/bitstream/123456789/20372/5/Chapter%20I.pdf

Sekolah Tinggi Ilmu Statistik. (2010). *Pedoman Penyusunan Skripsi Jurusan Komputasi Statistik Sekolah Tinggi Ilmu Statistik (Edisi Keempat, Revisi 1/2014)*. Jakarta: Sekolah Tinggi Ilmu Statistik.

Yogantara, I. (2010). *Pengembangan Sistem Aplikasi Pengelolaan Resume Berita Media Cetak dan Online secara Terpadu* [Skripsi]. Jakarta: Sekolah Tinggi Ilmu Statistik.

Young, M. J. (2001). *Step by Step XML*. Terjemahan oleh Imam Mustaqim. Jakarta: PT Elex Media Komputindo.

http://en.wikipedia.org/wiki/Apache_POI/, diakses pada 30 Agustus 2014, 15:39 WIB.

http://jsoup.org/, diakses pada 30 Agustus 2014, 15:39 WIB.