Prosiding Seminar Ilmiah Nasional Komputer dan Sistem Intelijen (KOMMIT 2014) Vol. 8 Oktober 2014
Universitas Gunadarma – Depok – 14 – 15 Oktober 2014 ISSN : 2302-3740

KLASIFIKASI BERITA BERDASARKAN PENDEKATAN SEMANTIK

***Dimas Bagus Prasetyo*[1]**
***Freddy Arviando*[2],**
***Muhammad Farhan Mubarak*[3]**
***Sandy Febriant Eka Putra*[4]**

[1,2,3,4]*Jurusan Teknik Informatika, Fakultas Teknologi Industri, Universitas Gunadarma*

Abstrak

Berita elektronik telah menjadi semakin popular sejak dimulainya perkembangan internet. Melalui internet, berita elektronik dikemas sedemikian rupa sehingga mampu mengalirkan informasi secara up-to-date kepada masyarakat. Namun, hal ini berdampak pada ketersediaan berita elektronik yang terlalu melimpah. Terlalu berlimpahnya jumlah item berita elektronik akan mengakibatkan sulitnya mencari berita yang relevan yang diinginkan pengguna. Salah satu solusi untuk mengatasi masalah ini adalah dengan menggunakan sistem yang mampu mengklasifikasikan berita secara relevan. Paper ini menganalisis proses pengklasifikasian berita dengan menggunakan pendekatan semantik. Pendekatan semantik mampu menganalisis informasi yang relevan dalam lingkup domain ontologi. Terdapat dua tahapan proses yang akan dianalisis, yaitu proses klasifikasi berita dan yang terdapat pada proses knowledge base.

Kata Kunci : *Berita, Klasifikasi, Knowledge Base, Semantik*

PENDAHULUAN

Berita berperan dalam memberikan informasi yang aktual mengenai kejadian penting yang sedang berlangsung. Saat ini, dengan dukungan teknologi yang semakin canggih, terutama teknologi informasi, proses pendistribusian informasi telah mengalami perubahan yang signifikan. Para pelaku bisnis yang bergerak di bidang jurnalistik mengubah cara penyampaian berita, dari era media cetak menuju era digital. Di era digital, memungkinkan para pengusaha mengemas informasi dalam bentuk berita elektronik (digital).

Web merupakan salah satu *platform* yang paling popular dalam mendistribusikan berita elektronik. Terdapat beberapa alasan mengapa web menjadi populer, seperti misalnya mengurangi biaya untuk distribusi dan akses berita, ketersediaan web pada banyak *platform browser*, penyampaian informasi yang mampu menjangkau seluruh dunia, dan mengurangi waktu yang dibutuhkan untuk publikasi berita.

Sayangnya, web juga merupakan penyebab dari salah satu masalah paling serius: jumlah berita yang dirilis setiap harinya melalui portal berita menjadi sangat banyak. Hal ini berdampak pada ketersediaan berita yang jumlahnya terlalu melimpah. Karena item berita yang tersedia jumlahnya melimpah, sulit untuk mengekstrak hanya item berita yang

relevan. Salah satu solusi yang dapat digunakan untuk mengatasi masalah ini adalah dengan menggunakan sistem yang mampu mengklasifikasikan berita berdasarkan pendekatan semantik.

Paper ini menganalisis proses pengklasifikasian berita dengan menggunakan pendekatan semantik. Pendekatan semantik digunakan untuk mengambil item berita yang saling berhubungan baik secara langsung maupun tidak langsung, dengan konsep dari domain ontologi. Ide utama dari pendekatan semantik diadopsi dari *knowledge base*, yang mendasari hubungan antara konsep-konsep ini. Ketersediaan konsep ini memungkinkan untuk menghasilkan item berita yang relevan kepada pengguna.

Landasan Teori

Di dalam proses pengklasifikasian berita dibutuhkan beberapa tahapan, yaitu pembuatan *knowledge base*, proses klasfikasi, ontologi berita, dan peng-*update*-an *knowledge base*. Sebelum membahas tentang proses tersebut, terlebih dahulu dijabarkan mengenai landasan teori yang digunakan.

Pendekatan Semantik

Model semantik terdiri dari jaringan konsep-konsep (*network of concepts*) dan hubungan antara konsep-konsep tersebut. Konsep adalah tema atau topik tertentu yang menjadi perhatian pengguna. Model semantik memungkinkan pengguna untuk mengidentifikasi pola dan tren dalam informasi dan juga menemukan hubungan antar potongan-potongan informasi yang berbeda. Konsep dan hubungan antar konsep-konsep sering dikenal sebagai Ontologi [8].

Semantic web diperkenalkan dengan penggunaan pendekatan semantik untuk memecahkan permasalahan keragaman informasi. Ontologi telah menjadi alat yang menarik dan menantang untuk pendekatan semantik. Interoperabilitas semantik dicapai menurut hubungan antar terminologi ke ontologi yang berseberangan, seperti menggunakan sinonim, hiponim dan hipernim [1].

Ontologi

Ontologi merupakan representasi formal pengetahuan dari kumpulan informasi (konsep) dalam suatu domain, dan hubungan antar konsepnya. Pada dasarnya informasi disimpan dalam ontologi dan dibagi menjadi tiga kategori berbeda: pertama, *knowledge base* yang mengandung konsep yang relevan untuk domain berita yang akan dianalisis, di dalam *knowledge base* juga disimpan aturan dari ontologi untuk memperbaharui informasi yang baru. Kedua, item berita yang telah diklasifikasikan, dan ketiga adalah hubungan antara item berita dan konsep yang terdapat di dalam *knowledge base.*

Knowledge Base

Knowledge base adalah teknologi yang digunakan untuk menyimpan informasi yang kompleks baik yang terstruktur maupun tidak terstruktur yang nantinya akan digunakan oleh sistem komputer. *Knowledge base* ini disimpan dalam ontologi yang berisi tentang informasi yang berkaitan dengan domain berita tertentu. Di dalam *knowledge base* terdapat konsep yang relevan dengan domain berita tertentu, sehingga item berita yang ingin diklasifikasikan dapat dianalisis kecocokannya dengan domain berita tertentu.

Informasi dalam knowledge base disimpan dalam format OWL (Web Ontology Language), yaitu bahasa yang secara formal dapat menggambarkan makna dari suatu konsep. Dengan format tersebut, semantik dari informasi tersebut tidak hanya dapat dimengerti oleh komputer tapi juga memungkinkan adanya pertukaran data antar komputer. Ketika data dianotasikan dengan format OWL, komputer dapat dengan mudah mengintegrasikan data dengan informasi lain. Jika semua item berita dianotasikan dengan benar, akan lebih mudah untuk mengekstrak item berita yang relevan dengan domain berita tertentu.

Penyimpanan konsep di *knowledge base* memungkinkan adanya proses saling berbagi (*sharing*) informasi dalam ontologi, mendukung kemudahan penambahan informasi (*extensibility*) dan juga bisa digunakan di berbagai lingkungan perangkat lunak. Penyimpanan di ontologi juga memudahkan adanya pertukaran data antara komputer. Komputer lain dapat menggunakan *knowledge base* kita dan juga kita dapat menggunakan *knowledge base* domain berita yang lain , sehingga kita bisa memanfaatkan *knowledge base*-nya untuk mengklasifikasikan berita dan menambah ontologi berita.
Untuk membuat pertukaran informasi menjadi sesederhana mungkin maka tidak ada pembatasan khusus selama informasi disimpan dalam format OWL. Setiap konsep dalam ontologi memiliki properti sinonim yang didefinisikan untuk mendukung proses klasifikasi. Selain itu juga terdapat properti WordNet yang berfungsi menyimpan URI (Uniform Resource Identifier) dari WordNet agar dapat mengambil lebih banyak lagi sinonim yang terdapat pada WordNet. WordNet juga menyediakan hiponim dan hipernim untuk *synsets* yang dapat berguna pada beberapa kasus dalam proses klasifikasi [4].

Ontologi Berita

Ontologi berita berfungsi untuk menguhubungkan antara konsep *knowledge base* dengan item berita itu sendiri. Ontologi berita menyimpan item berita dengan semua informasi yang relevan, termasuk hyperlink, waktu rilis, dan sumber. Proses klasifikasi menghubungkan item berita dengan konsep di *knowledge base* dan sebaliknya. Hubungan ini disimpan dalam domain ontologi. Dengan menggunakan hubungan antara konsep di *knowledge base* dan item berita, kita dapat menyediakan personalisasi berita. Dengan menggunakan dua ontologi yaitu ontologi *knowledge base* dan ontologi berita, kita sudah dapat memulai proses klasifikasi dan mengisi ontologi berita dengan item berita yang terhubung dengan konsep di ontologi *knowledge base* [4].

Wordnet

Wordnet adalah sebuah basis data yang berisi jaringan semantik untuk bahasa Inggris yang dikembangkan oleh Princeton University. Komponen utama dari wordnet berupa *synset*, yaitu sekumpulan sinonim yang dari suatu konsep (kata), beserta deskripsi makna dari konsep tersebut. *Synset* berbeda dengan "kata" (words), tapi merupakan sekumpulan makna kata yang bersinonim [6]. *Synset* juga dihubungkan dengan berbagai bentuk relasi seperti *hypernym* (adalah jenis dari), *meronym* (adalah bagian dari), *antonym* (adalah lawan dari) dan sebagainya [7].

Pengklasifikasian Berita

Proses klasifikasi pada dasarnya terdiri dua tahap, yaitu: mencari nama dan

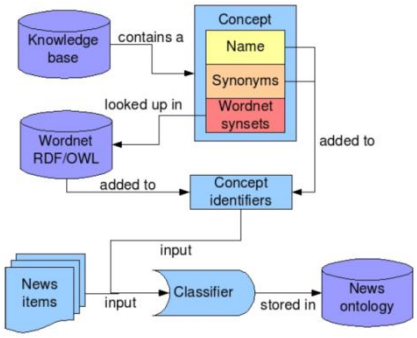
sinonim dari konsep yang didefinisikan di dalam ontologi, dan mencari kata-kata yang menyatakan konsep dengan bantuan aplikasi WordNet. Proses klasifikasi merupakan knowledge base yang sentris karna konsep diambil dari knowledge base dan akan dicocokkan secara langsung dengan setiap item berita. Pendekatan sentris item berita berarti untuk setiap konsep pada item berita dalam ontologi akan dicocokkan. Pendekatan sentris dipilih karena memiliki kinerja yang lebih baik pada kombinasi dengan mengambil sinonim dan hiponim dari WordNet, yang harus dilakukan setiap kali sebuah konsep diambil dari knowledge base. Sebelum memulai proses pengklasifikasian, terlebih dahulu dilakukan proses *Stopwords* dan juga *Stemming* [4].

Stopwords dan Stemming

Sebelum melakukan proses stemming, dilakukan terlebih dahulu pembuangan kata-kata yang tidak penting atau disebut *stopwords*. Kata-kata tidak penting misalnya kata sambung seperti dan, atau, jika, dll. Selain itu perlu juga pengeliminasian tanda baca. Setelah itu barulah dilakukan proses *stemming*, adalah pengolahan kata untuk mendapatkan kata dasar dari sebuah kata dengan mereduksi imbuhan dari kata tersebut dengan asumsi bahwa kata-kata tersebut memiliki makna yang sama pula, misalnya kata *connection, connective, connected* memiliki kata dasar *connect*. Pada dasarnya jenis imbuhan kata dalam bahasa Indonesia terdiri dari 3 imbuhan yaitu awalan, akhiran dan sisipan. Dari proses *stemming* ini nantinya akan dihasilkan indeks kata-kata yang dipertimbangkan untuk mewakili sebuah berita [5].

HASIL DAN PEMBAHASAN

Sejumlah item berita dirilis setiap harinya oleh portal berita, namun tidak semuanya sesuai dengan domain berita yang diinginkan. Maka dari itu, diperlukan pengklasifikasian sejumlah berita yang banyak tersebut sehingga hanya berita yang relevan dengan domain berita spesifik yang diinginkan saja yang nantinya akan muncul. Proses pengklasifikasian berita tersebut dapat digambarkan sebagai berikut :



Gambar 1. Proses Klasifikasi [4].

Untuk dapat mengklasifikasikan berbagai macam item berita, *knowledge base* yang berdasarkan domain berita tertentu harus dibuat terlebih dahulu. Dari *knowledge base* yang dibuat tersebut nantinya kita dapat mengklasifikasikan item berita sesuai ke dalam domain berita tertentu dengan melalui proses *concept matching*. Nantinya *knowledge base* akan terus di-*update* berdasarkan berita yang masuk ataupun berdasarkan event tertentu.

Proses pada Ontologi

Knowledge base berisi konsep-konsep tentang domain berita tertentu. konsep tersebut memiliki *synonym property* untuk mendapatkan kata-kata yang memiliki makna serupa. Konsep juga memiliki WordNet *property* sehingga sistem dapat mencari lebih banyak sinonim, hiponim dan hipernim yang berhubungan dengan konsep yang ada di Wordnet.

Dengan menggunakan kueri, pengguna dapat memilih konsep dari *knowledge base* untuk mendapatkan berita yang diinginkan. Konsep tersebut lalu ditambahkan ke *Concept Identifier*. Sejumlah item berita yang nantinya akan dicocokkan dengan *concept* yang dipilih pengguna untuk disaring sehingga hanya berita yang relevan saja yang bisa diterima oleh sistem.

Item berita yang telah diklasifikasi dan hubungan antara *concept* di *knowledge base* dengan item berita yang telah diklasifikasi selanjutnya akan disimpan dalam ontologi berita. Lalu dari proses tersebut kita dapat menambah knowledge base kita berdasarkan berita yang sudah diklasifikasi tersebut. *Knowledge base* juga bisa ditambahkan berdasarkan pada *event-event* tertentu.

Studi Kasus

Pada suatu ketika ditentukan domain berita yang diinginkan, yaitu teknologi. Untuk mengetahui apakah berita tersebut tergolong ke dalam klasifikasi teknologi atau tidak, maka digunakan metode pemrosesan pada ontologi.

Gambar 2. Contoh paragraf berita teknologi, detik.com

Sebelum melakukan pengklasifikasian, terlebih dahulu melakukan proses *Stopwords* pada item berita, yaitu menghilangkan kata sambung. Kemudian, dilakukan proses Stemming (menghilangkan iimbuhan), sebagai contoh "mengunduh", maka kata dasar yang diperolah ialah "unduh".

Maskapai Emirates meluncurkan aplikasi pengguna iPad. Traveler memesan tiket pesawat Emirates mudah aplikasi, traveler mengunduh boarding pass dicetak. Maskapai Emirates meluncurkan aplikasi pengguna iPad. Traveler memesan tiket pesawat Emirates mudah aplikasi, traveler mengunduh boarding pass dicetak.

Gambar 3. Contoh hasil proses *Stopwords*.

[meluncurkan : luncur, memesan : pesan, mengunduh : unduh]

Gambar 4. Contoh hasil proses *Stemming*.

Setelah itu baru kata tersebut dicocokkan dengan kata-kata yang berada di *Knowledge Base*, bisa juga sinonim dari

[iPad, aplikasi, unduh, aplikasi, iPad, aplikasi, unduh]

kata yang dicari, misalkan kata "unduh" diperoleh dari sinonim dari kata "*download*". Jika kata yang dicari tersebut tidak diperoleh pada sinonimnya, maka dibutuhkan aplikasi tambahan yaitu WordNet untuk mencari kata yang mempunyai relasi dengan kata tersebut. Relasi ini bisa berupa hiponim ataupun hipernim dari kata yang dicari. Sebagai contoh, kata yang akan dicari ialah "handphone". Jika handphone tidak ditemukan sinonimnya di konsep pada *knowledge base*, maka WordNet akan membantunya. Misalkan, di WordNet terdapat kata "elektronik". Dikarenakan handphone merupakan hiponim dari suatu hipernim elektronik, maka kata tersebut bisa disimpan ke dalam *Concept Identifier*.

Jakarta - Maskapai Emirates meluncurkan aplikasi untuk pengguna iPad. Traveler kini bisa memesan tiket **pesawat** Emirates secara lebih mudah lewat aplikasi, traveler juga bisa mengunduh boarding pass yang bisa langsung dicetak.

Gambar 5. Gambaran isi dari *Concept Identifier*.

Selanjutnya, pengguna melakukan kueri dengan memilih domain berita bahkan konsep (informasi) yang dibutuhkan pengguna. Diasumsikan item berita yang diinginkan ialah yang berkaitan dengan teknologi, finansial, dan politik. Semua kata yang sudah ditampung pada *Concept Identifier* dan juga item berita yang telah ditentukan akan dijadikan sebagai inputan untuk dilakukan proses pengklasifikasian

berita. Apabila banyaknya kata yang diperoleh sesuai dengan ketentuan pada proses klasifikasi dari suatu item berita, maka outputnya akan ditampung pada ontology berita yakni berisikan kumpulan berita yang telah selesai diklasifikasikan. Dalam kasus ini, sebagai contoh bahwa jumlah kata yang berkaitan dengan teknologi di dalam suatu berita diperoleh sebanyak 20 kata, dimana ketentuan dari klasifiksi dari item berita teknologi ialah minimal 10 dari 100 kata. Sehingga di dalam ontology berita, item berita tersebut diklasifikasikan sebagai berita teknologi.

SIMPULAN DAN SARAN

Dengan proses klasifikasi tersebut, berita elektronik akan diklasifikasikan dan disaring agar sesuai dengan domain berita tertentu. Untuk mendapatkan item berita yang sesuai dengan ketertarikannya, user dapat memilih konsep dari *knowledge base* dengan melakukan perintah kueri untuk mengekstrak sejumlah berita yang relevan dengan yang diinginkan oleh pengguna.

DAFTAR PUSTAKA

Wicaksana, I Wayan Simri. Banowosari, Lintang Yunia. Wulandari, Lily. Wirawan, Setia, "Pentingnya Peranan Bahasa dalam Interoperabilitas Informasi Berbasiskan Komputer karena Keragaman Semantik", 2005, diakses pada tanggal 11 April 2014.

Purwitasari, Diana. Maulana, Rizki Akbar. Shiddiqi, Ary Mazharuddin, "Framework Sistem Rekomendasi Berita Berbahasa Indonesia Berdaasarkan Pilihan Minat Baca Personal", 2012, diakses pada 12 April 2014.

IJntema, Wouter. Goossen, Frank. Frasincar, Flavius. Hogenboom, Frederik, "Ontology-Based News Recommendation", 2010 , diakses pada tanggal 14 April 2014.

Borsje, Jethro. Levering, Leonard. Frasincar, Flavius, "Hermes: a Semantic WebBased News Decision Support System", 2008, diakses pada tanggal 14 April 2014, Chen, Yen-Liang. Yu, Tung-Lin, "News Classification based on experts' work knowledge", 2011, diakses pada tanggal 15 April 2014

Maharani, Warih. Firdaus, Yanuar, "Analisis Semantic Similarity pada Item Based Recommender System", 2008, diakses pada tanggal 15 April 2014.

Wicaksana, I Wayan Simri, "Membandingkan Pendekatan *Latent Semantic* terhadap WordNet untuk *Semantic Similarity*", 2006, diakses pada tanggal 15 April 2014.

Ortega de Mues, Mariano. Espinoza Angelina, Rodriguez-Alvarez, Daniel, "Harmonization of Semantic Data Models of ElectricData Standards", 2011, diakses pada tanggal 24 April 2014.