

ANTARA: TESAURUS MULTIBAHASA BERBASIS KOMPONEN

Siti Rochimah, Alexander L. Romy

Jurusan Teknik Informatika,

Fakultas Teknologi Informasi, Institut Teknologi Sepuluh Nopember

Kampus ITS, Jl. Raya ITS, Sukolilo – Surabaya 60111, Telp. + 62 31 5939214, Fax. + 62 31 5913804

Email: siti@its-sby.edu, romyinf99@plasa.com

ABSTRAK

Tesaurus merupakan kosakata yang terkontrol yang telah banyak dipakai untuk pengindeksan dan pemahaman kata-kata. Pemakaian tesaurus memerlukan kelengkapan dan kecepatan oleh karena itu banyak dibuat program tesaurus versi offline dan online. Tugas akhir ini bertujuan membangun program tesaurus multibahasa manual yang lengkap terutama dalam hal pemasukan elemen-elemen tesaurus yang telah menjadi standard dalam pembangunan tesaurus yang bernama AntarA. Pembangunan tesaurus akan difokuskan pada arsitektur agar dapat dibuat pemisahan antara komponen utama dan bagian antar muka. Hal lain yang ikut diperhitungkan adalah menyangkut kegunaan ulang (reusability) dari komponen yang dibuat dan kecepatan akses pengguna dalam berbagai proses yang dilakukan program.

Metodologi yang dipakai dalam penelitian ini adalah dengan mempelajari dan merancang basis data yang mendukung elemen-elemen tesaurus yaitu term, relasi, definisi dan presentasi. Selanjutnya dikembangkan tiga modul utama yaitu modul pencarian, modul editor dan modul manajer basis data. Modul pencarian berfungsi untuk mempresentasikan tesaurus dengan bermacam-macam tipe tampilan, modul editor berfungsi untuk modifikasi term, relasi dan definisi, sedangkan modul manajer basis data dibuat dalam bentuk komponen. Untuk pengisian data awal, dipakai data dari tesaurus UNESCO. Secara keseluruhan AntarA diimplementasikan dalam lingkungan Visual Studio .NET dengan sistem operasi Windows XP.

Aplikasi telah diuji coba pada sistem sesuai dengan spesifikasi dan kebutuhan berdasarkan skenario yang dirancang sendiri untuk setiap modul AntarA. Hasil uji coba menunjukkan bahwa AntarA dapat diimplementasikan dengan baik, dengan demikian AntarA dapat dipakai sebagai tesaurus multibahasa digital.

Kata Kunci: tesaurus, tesaurus multibahasa, tesaurus digital, presentasi tesaurus, komponen, relasi kata, relasi term, bahasa, kamus

1. PENDAHULUAN

Konsep tesaurus tidak berubah dalam beberapa puluh tahun terakhir yaitu tesaurus merupakan bahasa pengindeksan yang dikembangkan institusi individual, dengan fungsi untuk mengoptimalkan pengindeksan dan pengambilan informasi sesuai kebutuhan user. Dalam 10 tahun terakhir, muncul daya tarik baru terhadap tesaurus yang terlihat dengan semakin banyaknya tesaurus digital yang dikembangkan. Daya tarik tersebut disebabkan beberapa faktor antara lain semakin murah harga komputer mikro, perkembangan internet dan meningkatnya kebutuhan akan tesaurus multilingual untuk perolehan informasi dalam berbagai bahasa. Menurut harga komputer mikro menyebabkan semakin meningkatnya pemakaian komputer pribadi dan program-program digital termasuk tesaurus multilingual dan jika dipadukan dengan internet maka dapat memberikan akses yang lebih luas terhadap pengguna [2].

Dalam tugas akhir ini, tesaurus multibahasa digital yang akan dikembangkan diberi nama **AntarA**, yang merepresentasikan jembatan antar

bahasa. Fitur unggulan AntarA adalah kelengkapan struktur tesaurus dalam basis data dan penambahan data oleh pengguna secara manual atau lewat XML. Pembangunan tesaurus akan menggunakan komponen yang dibuat sendiri untuk mempermudah pengembangan lebih lanjut di masa depan.

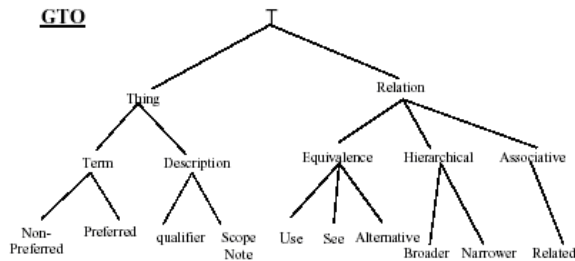
2. KONSEP TESAURUS

Tesaurus merupakan suatu kosakata terkontrol yang diatur secara terurut dimana relasi ekivalensi, hirarki dan asosiasi antar term ditampilkan secara eksplisit dan diidentifikasi dengan indikator relasi standard [4].

Tesaurus memiliki fungsi berdasarkan fungsionalitasnya yaitu:

- Tesaurus *stand-alone*
Berfungsi untuk membantu pengguna memahami arti suatu term beserta relasinya, membantu mencari variasi term dan mencari translasi term.
- Tesaurus terintegrasi
Memiliki dua fungsi utama yaitu pengambilan informasi dan pengindeksan.

Tesaurus memiliki struktur yang berbentuk pohon. Struktur tesaurus secara umum (*General Thesaurus Ontology*) atau GTO dapat terlihat sebagai berikut ini.



Gambar 1. Struktur Tesaurus / GTO

Dari gambar GTO dapat terlihat beberapa elemen tesaurus yaitu:

- Term formal (*descriptor*)
- Term non-formal (*non-descriptor*)
- Deskripsi term yaitu definisi dan *scope note*
- Relasi antar term
- Relasi lain yang dapat ditambahkan sendiri misalnya relasi translasi antar bahasa (TR)

Elemen lain yang penting adalah presentasi tesaurus. Berdasarkan [1], [4], [2], [5] ada sejumlah presentasi tesaurus yaitu:

- Alfabetik
- Hirarki
- Sistematis
- Ekspansi
- Rotasi

Sebagai program tesaurus, AntarA harus dapat memenuhi beberapa persyaratan query tesaurus [4], yaitu:

- Mencari suatu node pada tree
- Mencari path dari root ke suatu node pada tree
- Mencari suatu subtree
- Mencari suatu set node dengan relasi asosiasi

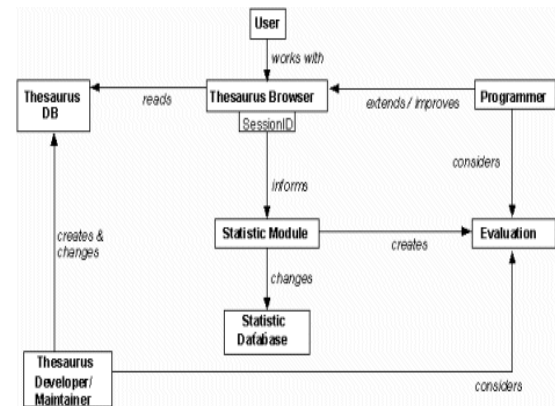
Suatu program tesaurus memiliki interaksi dengan lingkungan sekitarnya yang dapat digambarkan seperti gambar 2. [3]

Dari sejumlah elemen lingkungan program tesaurus tersebut, pembangunan AntarA akan difokuskan pada elemen pengguna, basis data tesaurus dan browser tesaurus.

3. DESKRIPSI ANTARA

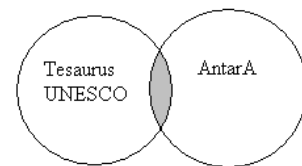
AntarA adalah tesaurus multibahasa offline yang bersifat *stand-alone*, manual, *free*, khusus untuk platform Windows yang telah terinstall dotNet framework versi 1 ke atas. Multibahasa berarti AntarA adalah tesaurus yang memiliki lebih dari dua bahasa dengan satu bahasa dominan yaitu bahasa Inggris. *Offline* berarti AntarA hanya dapat dipakai sebagai aplikasi *desktop* tanpa perlu koneksi internet. *Stand-alone* berarti AntarA tidak dapat diintegrasikan dengan perangkat lunak lain misalnya aplikasi *word*

processing atau *spreadsheet*. Manual berarti pemasukan basis data AntarA dilakukan secara manual. *Free* berarti paket aplikasi AntarA dapat di-*download* dari situs AntarA dengan gratis untuk kemudian diinstall pada komputer klien.



Gambar 2. Lingkungan Program Tesaurus Perbandingan Antara Dengan Tesaurus Unesco

Pada dasarnya tesaurus UNESCO merupakan bank data yang dipakai AntarA untuk mengisi data awal tesaurus. Tesaurus UNESCO dipilih sebagai bank data awal karena memiliki struktur yang mirip dengan AntarA. Kedua tesaurus memiliki hubungan saling beririsan seperti berikut ini.



Gambar 3. Hubungan Tesaurus UNESCO dengan AntarA

Berikut perbandingan fungsionalitas AntarA dan tesaurus UNESCO.

Tabel 1. Perbandingan AntarA dan Tesaurus UNESCO

Fungsionalitas	AntarA	UNESCO
Tipe relasi	Hirarki, asosiasi, ekivalensi inklusi dan translasi	Hirarki, asosiasi, ekivalensi inklusi dan inter bahasa
Pemasukan data	Dilakukan melalui editor	Tidak dapat dilakukan
Modifikasi data	Dilakukan melalui editor	Tidak dapat dilakukan
Penghapusan data	Dilakukan melalui editor	Tidak dapat dilakukan
Tampilan ke layar	Ya, secara offline	Ya, secara online
Keluaran ke printer	Tidak ada	Tidak ada
Keluaran ke file	Ya, dalam bentuk XML	Tidak ada

Hal lain yang perlu diketahui adalah tesaurus UNESCO memiliki dua versi yaitu versi online yang bebas dipakai pengguna dan versi offline yang harus dibeli. AntarA hanya memiliki satu versi offline yang dapat didownload secara bebas.

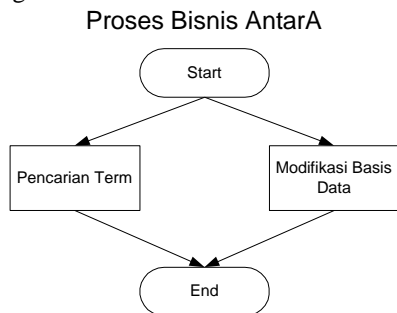
4. ANALISIS KEBUTUHAN ANTARA

4.1. PROSES BISNIS ANTARA

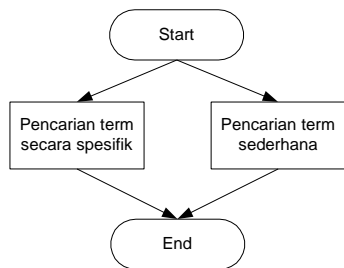
Proses bisnis AntarA terdiri dari dua bagian besar yaitu bagian modifikasi basis data dan pencarian data. Modifikasi basis data dibagi lagi menjadi dua yaitu modifikasi lewat form dan lewat XML, sedangkan pencarian data juga dibagi dua yaitu pencarian secara spesifik (*expert search*) dan pencarian secara sederhana (*simple search*).

Pencarian term merupakan fungsi utama dari AntarA sedangkan modifikasi basis data digunakan untuk penambahan data oleh ahli bahasa atau pengguna yang ingin menambahkan sendiri term-term yang belum terdapat dalam basis data.

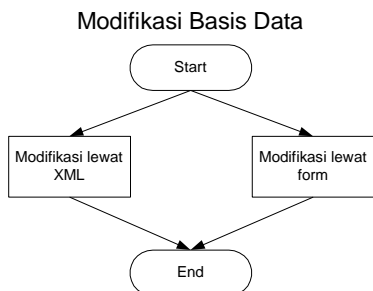
Secara garis besar hal itu terlihat pada pemodelan proses bisnis AntarA seperti pada gambar-gambar berikut ini.



Gambar 4. Proses bisnis AntarA level 0 Pencarian Term



Gambar 5. Proses bisnis AntarA level 1 untuk pencarian term



Gambar 6. Proses bisnis AntarA level 1 untuk modifikasi basis data

4.2. DIAGRAM ALIR ANTARA

Diagram alir AntarA dibuat berdasarkan diagram proses bisnis pada bagian sebelumnya. Oleh karena itu, diagram alir (*flowchart*) AntarA juga dapat dibagi menjadi dua bagian yaitu pencarian term dan modifikasi basis data. Diagram alir AntarA selengkapnya dapat dilihat pada Gambar 7.

4.3. PEMODELAN USE CASE ANTARA

Pada diagram use case AntarA hanya terdapat satu aktor yaitu pengguna tesaurus secara umum. Aktor dalam sistem AntarA dapat dibagi menjadi dua yaitu pengguna ahli (*expert user*) dan pengguna umum (*common user*). Pengguna ahli adalah pengguna yang memiliki pengetahuan dalam bidang bahasa dan biasanya memerlukan informasi yang lebih mendalam mengenai hasil pencarian term. Pengguna ahli juga memiliki keahlian dalam memodifikasi data tesaurus dengan tepat. Pengguna umum adalah pengguna tesaurus yang tidak memiliki pengetahuan bahasa secara mendalam dan biasanya hanya memerlukan informasi yang seperlunya mengenai hasil pencarian term. Pengguna umum tidak memiliki keahlian dalam modifikasi data tesaurus. Meskipun secara akademis berbeda, pengguna umum dan pengguna ahli memiliki hak yang sama dalam pemakaian AntarA tanpa ada pembatasan atau pembedaan hak.

Diagram use case AntarA dapat dilihat pada Gambar 8. Masing-masing use case akan terdiri atas sub-sub use case yang lebih terinci.

4.4. PEMODELAN DIAGRAM SEKUENS ANTARA

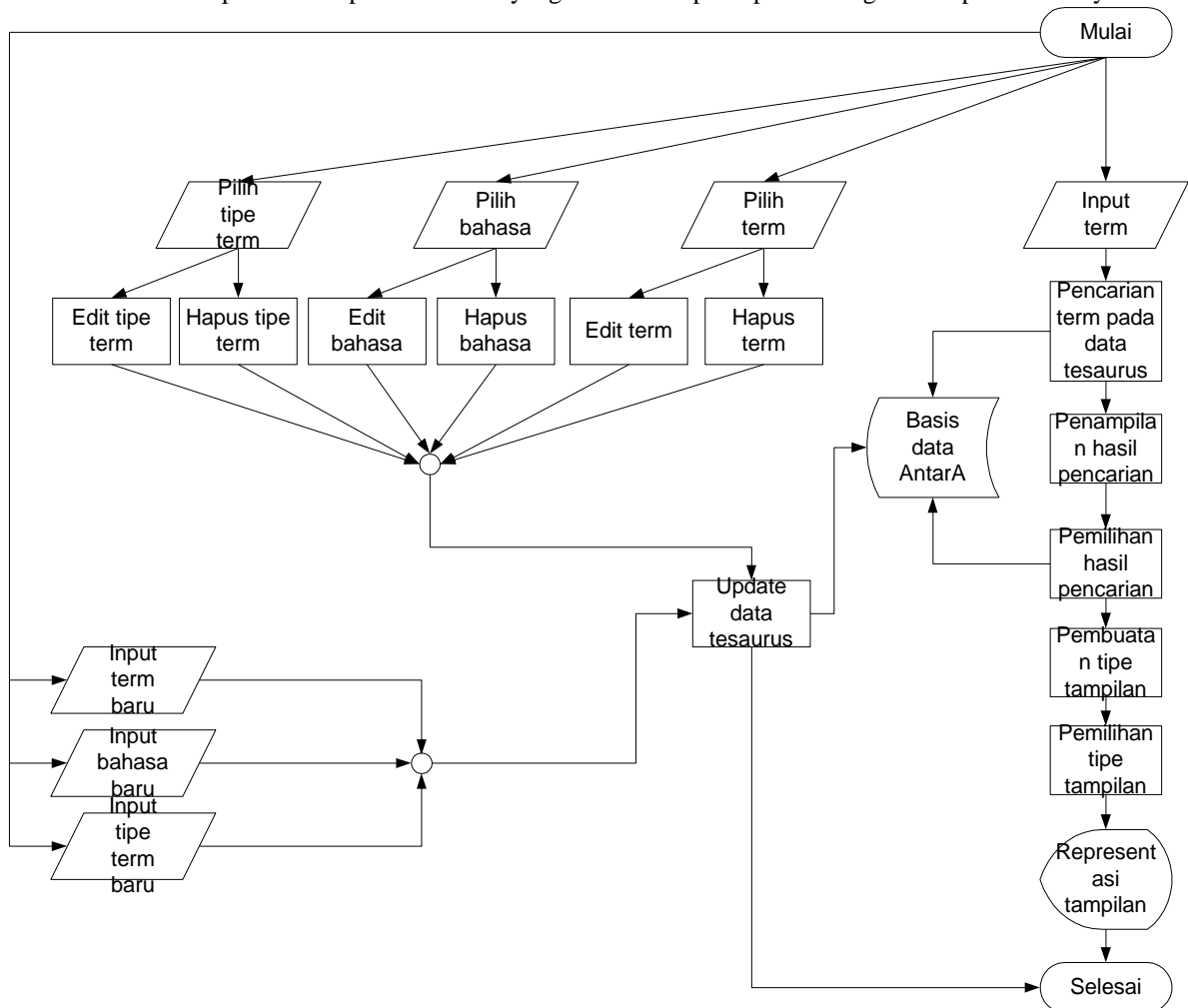
Sequence diagram merupakan diagram yang menjelaskan operasi yang dilakukan sistem berdasarkan urutan waktu untuk setiap realisasi use case. Berikut akan dijelaskan salah satunya (Gambar 9).

4.5. PEMODELAN KELAS ANTARA

Berdasarkan proses bisnis, pemodelan kelas-kelas AntarA dapat dikelompokkan berdasarkan sub sistem-sub sistem AntarA, yaitu:

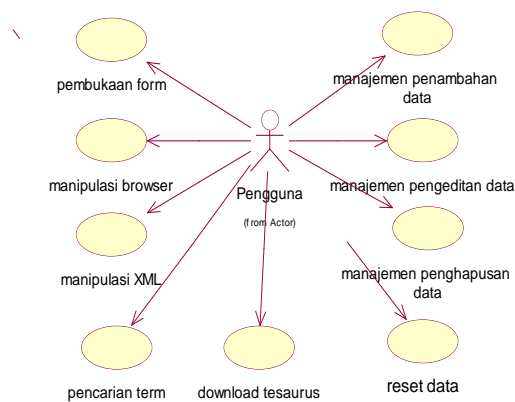
- AntarAMain
Terdiri atas kelas-kelas tampilan menu utama AntarA
- AntarAEditor
Terdiri atas kelas-kelas tampilan browser dan sebuah kelas manipulasi browser
- AntarASearch
Terdiri atas kelas-kelas tampilan pencarian dan sebuah kelas pembentukan hasil pencarian
- AntarADB
Terdiri atas kelas-kelas yang merupakan kontainer tabel-tabel dari basis data. AntarADB dapat dianggap sebagai manajer basis data.

AntarADB merupakan komponen utama yang dapat dipakai ulang untuk aplikasi lainnya.

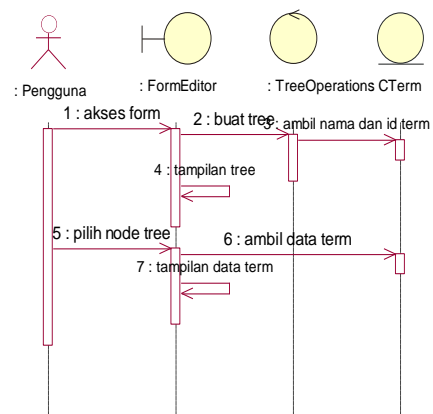


Gambar 7. Diagram Alir AntarA

Diagram kelas AntarA secara umum dapat dilihat pada Gambar 10. Untuk setiap subsistem dijabarkan lagi diagram kelasnya seperti terlihat pada Gambar 11.



Gambar 8. Diagram Use Case AntarA



Gambar 9. Diagram Sekuens untuk Use Case Browsing Tesaurus (Manipulasi Browser)

4.6. PEMODELAN BASIS DATA ANTARA

Basis data AntarA dirancang untuk kemudahan pengembangan tesaurus di masa depan khususnya dalam penambahan term dan jenis bahasa. Model fisik basis data AntarA dapat dilihat pada Gambar 12.

Dari pemodelan pada Gambar 12 terdapat beberapa hal penting yaitu:

- Suatu term dapat terdiri atas *field*, *microthesauri*, *top term* atau *narrower term* yang secara hirarki saling berhubungan
- Suatu term dapat memiliki sejumlah relasi dengan term lain
- Translasi bahasa juga merupakan relasi
- Definisi term tergantung dari tipe term

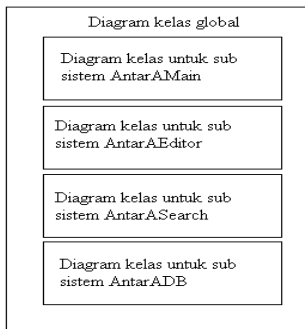
5. PERANCANGAN ANTARA

5.1. ARSITEKTUR ANTARA

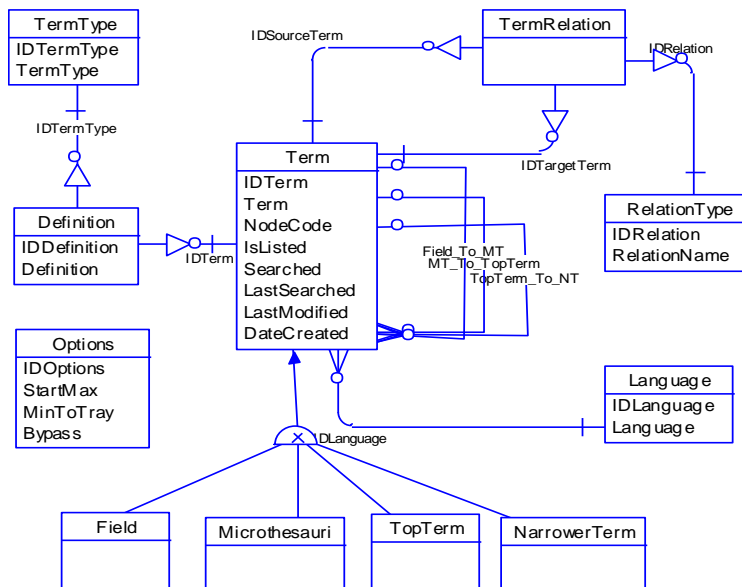
Arsitektur AntarA terdiri atas tiga layer yaitu:

- Layer antar muka (*graphical user interface / GUI*)
- Layer tesaurus
- Layer akses basis data

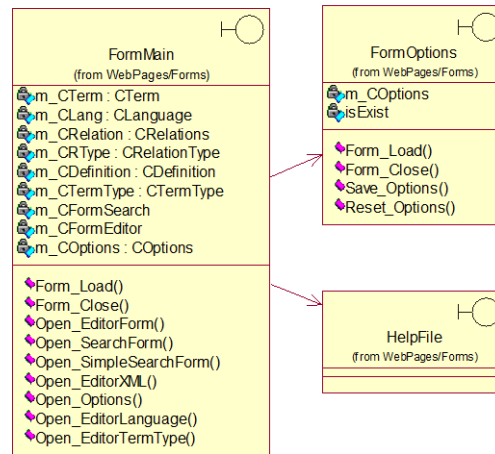
Arsitekturnya dapat dilihat pada Gambar 13.



Gambar 10. Diagram Kelas AntarA secara Global



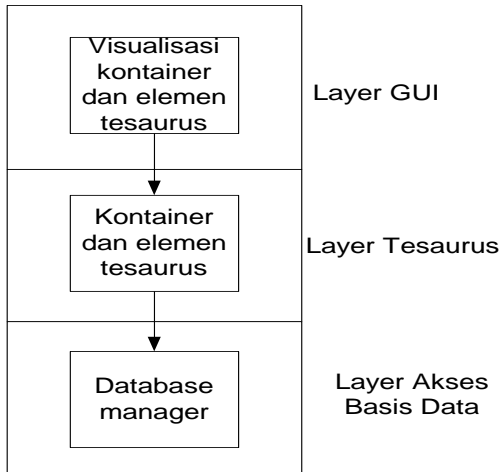
Gambar 12. Model Fisik Basis Data AntarA



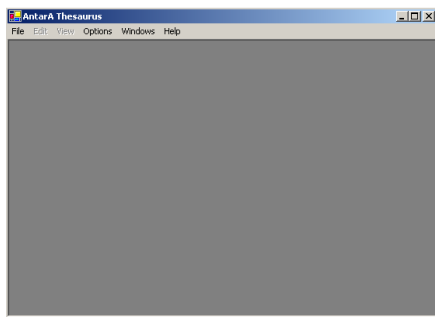
Gambar 11. Diagram Kelas untuk AntarAMain

5.2. PERANCANGAN ANTARMUKA

Sub Sistem antarmuka terdiri atas form-form yang bersifat interaktif bagi pengguna. Yang merupakan Sub Sistem antarmuka adalah Sub Sistem AntarAMain, AntarAEditor dan AntarASearch. Gambar 14 s/d 19 memperlihatkan beberapa diantaranya.

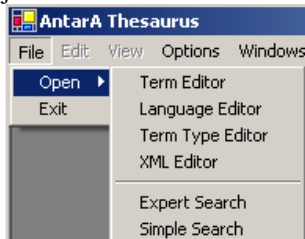


Gambar 13. Arsitektur AntarA

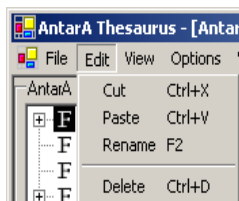


Gambar 14. Tampilan FormMain

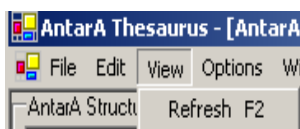
FormMain merupakan form utama yang pertama kali diakses oleh pengguna. Form ini memiliki sejumlah menu antara lain:



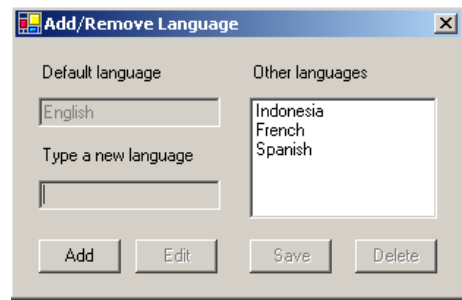
Gambar 15. Menu File pada FormMain



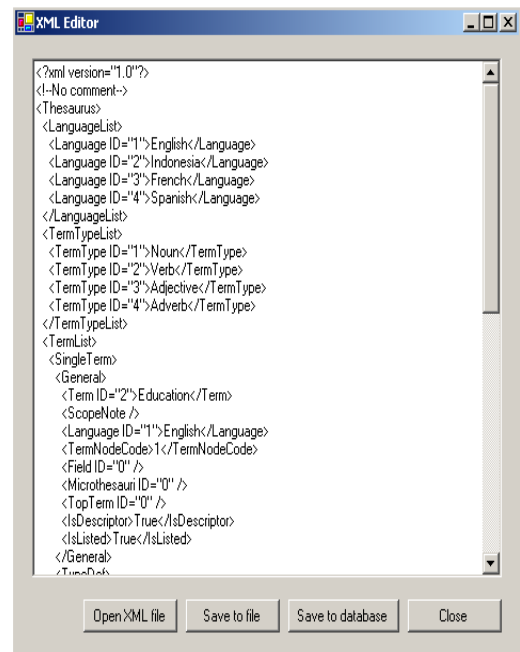
Gambar 16. Menu Edit pada FormMain



Gambar 17. Menu View pada FormMain



Gambar 18. Tampilan Form Pengeditan Bahasa (Form Editor)



Gambar 19. Tampilan Form Pwngeditan XML (Form Editor)

6. IMPLEMENTASI

Bahasa pemrograman yang digunakan untuk membuat perangkat lunak ini adalah Visual C# pada sistem operasi Windows XP yang telah diinstal dengan dotNet framework versi 1.06. Sistem operasi lain yang kompatibel adalah Windows 2000 dan NT.

Implementasi AntarA difokuskan pada dua modul (implementasi dari sub sistem) yaitu modul AntarAEditor dan AntarASearch.

Implementasi AntarAEditor dapat dibagi menjadi dua yaitu:

- Tampilan browser
- Manipulasi browser

Implementasi AntarASearch dapat dibagi menjadi enam sesuai dengan jumlah tampilan yang disediakan yaitu:

- Tampilan Alfabetik

Algoritma untuk tampilan ini adalah:

Jika bukan descriptor
Ambil relasi ekuivalensi ke descriptor

Tampilkan hasil
 Jika descriptor
 Ambil relasi inklusi
 Ambil semua relasi translasi
 Ambil semua relasi ekivalensi
 Ambil semua relasi Broader Term satu level ke atas
 Ambil semua relasi Narrower Term satu level ke bawah
 Ambil semua relasi asosiasi
 Tampilkan hasil

- Tampilan Hirarki
 Algoritma untuk tampilan ini adalah:

Cari lokasi term dalam struktur tesaurus
 Traverse parent dari term
 Simpan nama parent
 Tampilkan hasil

- Tampilan Sistematis
 Algoritma untuk tampilan ini adalah:

Cari lokasi term dalam struktur
 Cari lokasi field dari term
 Traverse field
 Ambil nama term pada level yang lebih rendah
 Ambil relasi term pada level yang lebih rendah
 Tampilkan hasil

- Tampilan Ekspansi
 Algoritma untuk tampilan ini adalah:

Ambil relasi BT satu level ke atas dan relasi lainnya
 Traverse tree mulai node term yang dicari
 Ambil relasi NT dan relasi lainnya
 Berhenti jika sudah tiga level ke bawah
 Hitung jarak semantik
 Ambil term ekspansi sesuai treshold
 Urutkan hasil perhitungan
 Tampilkan hasil

Algoritma untuk perhitungan jarak semantik adalah sebagai berikut:

Untuk setiap term hasil ekspansi
 Ambil string rute term
 Untuk setiap potongan path
 Jika BT tambahkan beban untuk BT
 Jika NT tambahkan beban untuk NT
 Jika RT tambahkan beban untuk RT
 Jika total beban <= threshold
 Masukkan term dalam Selected
 Urutkan Selected

- Tampilan Rotasi
 Algoritma untuk tampilan ini adalah :

Cari term secara eksak
 Cari term terdekat
 Untuk setiap kata dalam term yang dicari
 Untuk setiap term yang ditemukan
 Cari term yang memiliki kata tersebut di kontainer term
 Tampilkan hasil

7. UJICOBAN DAN EVALUASI

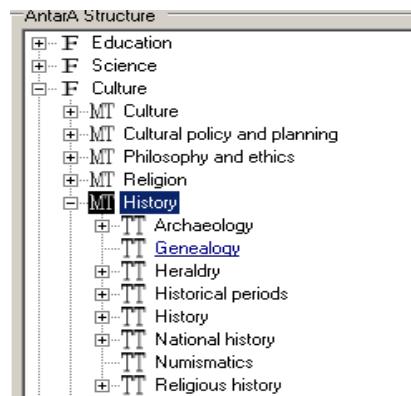
Uji coba AntarA dilakukan di rumah penulis. Spesifikasi komputer yang dipakai adalah prosesor Pentium Celeron 2 1.8 GHz dengan memori 256 MB. Parameter umum dalam uji coba ini adalah sebagai berikut:

- Kecepatan eksekusi setiap skenario
 - Konsistensi basis data terhadap perubahan data oleh pengguna
 - Kemudahan akses pengguna.
- Uji coba difokuskan pada dua modul utama yaitu AntarAEditor dan AntarASearch. Uji coba

pada modul AntarAEditor melibatkan penambahan, perubahan dan penghapusan berbagai elemen tesaurus. Berikut skenario-skenario untuk modul AntarAEditor:

Tabel 2. Skenario Ujicoba Modul AntarAEditor

Sub Modul	Skenario
Term Editor	<ol style="list-style-type: none"> 1. Menambah, mengupdate dan menghapus masing-masing descriptor 2. Menambah, mengupdate dan menghapus masing-masing non-descriptor 3. Memindahkan (cut & paste) descriptor secara bolak-balik 4. Menambah, mengupdate dan menghapus relasi asosiasi, ekivalensi dan translasi pada descriptor 5. Menambah, mengupdate dan menghapus definisi pada salah satu descriptor dan non-descriptor
XML Editor	<ol style="list-style-type: none"> 1. Membuka, mengedit dan menyimpan file XML yang benar 2. Membuka, mengedit dan menyimpan file XML yang salah
Language Editor	Menambah, mengedit dan menghapus bahasa
Term Type Editor	Menambah, mengedit dan menghapus tipe term

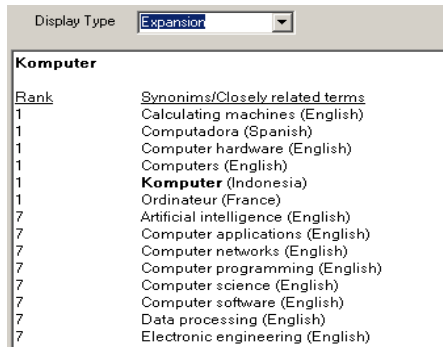


Gambar 20. Uji Coba Term Editor

Uji coba pada modul AntarASearch melibatkan pencarian berbagai term yang ada ataupun tidak ada dalam basis data serta juga diujicobakan setiap tampilan hasil pencarian. Berikut skenario-skenario untuk modul AntarASearch:

Tabel 3. Skenario Ujicoba Modul AntarASearch

Expert Search	<ol style="list-style-type: none"> 1. Mencari 5 term pada setiap field dengan tampilan yang berbeda 2. Mencari 5 term yang tidak ada dalam basis data 3. Mengetes tampilan untuk descriptor dan non-descriptor
Simple Search	<ol style="list-style-type: none"> 1. Mencari 10 term yang ada dalam basis data 2. Mencari 5 term yang tidak ada dalam basis data



Gambar 21. Uji Coba Expert Search

Evaluasi hasil uji coba untuk parameter-parameter umum adalah sebagai berikut:

- Kecepatan eksekusi setiap skenario
Kecepatan eksekusi untuk modul AntarAEditor dapat dilihat pada tabel berikut:

Tabel 4. Waktu Rata-rata untuk Eksekusi Setiap Skenario pada Modul AntarAEditor

Sub Modul	Skenario	Waktu rata-rata (detik)
Term Editor	1	24,875
	2	4,75
	3	16,833
	4	3,125
	5	0,5
XML Editor	1	10
Language Editor	1	14
Term Type Editor	1	10

Kecepatan eksekusi untuk modul AntarASearch bervariasi untuk setiap term tergantung dari banyaknya data dalam basis data. Masalah yang dihadapi adalah jika jumlah data dalam basis data bertambah maka kecepatan eksekusi akan berbanding terbalik.

- Konsistensi basis data
Parameter ini sudah terpenuhi karena setiap uji coba setiap modul berlangsung dengan baik. Konsistensi data dapat dijaga dengan membatasi akses pengguna pada setiap modul. Satu-satunya masalah yang mungkin terjadi adalah pada pemasukan data lewat XML. Pemasukan data dengan cara ini hanya direkomendasikan untuk pengguna ahli (*expert user*) saja karena dapat terjadi kesalahan pemasukan data yang berakibat pada kerusakan basis data.
- Kemudahan akses pengguna
Parameter ini diuji cobakan pada skenario download program, instalasi program, skenario modul Options dan skenario modul Help File. Secara keseluruhan semua uji coba parameter ini berlangsung dengan baik.

8. KESIMPULAN DAN SARAN

Beberapa kesimpulan yang dapat ditarik dari penelitian ini adalah sbb.:

1. AntarA dapat diimplementasikan dengan baik sebagai aplikasi Windows dengan persyaratan yang telah disebutkan.
2. AntarA telah memasukkan semua elemen tesaurus yang diperlukan sesuai standard pembuatan tesaurus yaitu term, relasi dan presentasi.
3. Komponen yang dibuat yaitu AntarADB.dll meliputi layer tesaurus dan layer basis data. Komponen tersebut ditujukan untuk pengembangan antar muka saja, pengembangan lebih lanjut harus mengikuti struktur dan arsitektur yang ada. Meskipun demikian, komponen tersebut belum diujicobakan untuk aplikasi yang lain.
4. AntarA merupakan tesaurus multibahasa tetapi pada dasarnya hanya ada satu bahasa yang dominan yaitu bahasa Inggris, bahasa lainnya hanya merupakan translasi yang bisa ditambahkan sebanyak mungkin.
5. Kecepatan proses AntarA dibatasi oleh jumlah data yang pada suatu saat akan mencapai titik maksimum yang belum dapat ditentukan. Untuk menentukan titik maksimum tersebut masih diperlukan penelitian lebih lanjut.
6. Struktur basis data tesaurus tidak rumit.
7. Masih diperlukan usaha lebih lanjut dalam hal pemasukan dan penambahan data tesaurus yang lebih lengkap. Kekurangan ini terjadi karena keterbatasan waktu dan tenaga.

9. DAFTAR PUSTAKA

1. Harith Alani, Christopher Jones dan Douglas Tudhope, "Associative and Spatial Relationships in Thesaurus-based Retrieval", School of Computing, University of Glamorgan, Wals, UK, Available from: <http://eprints.ecs.soton.ac.uk/archive/00005702>, 2001
2. Tim Craven, "Thesaurus Construction". Faculty of Information and Media Studies, University of Western Ontario, London, Ontario, Canada. Available from: <http://publish.uwo.ca/~craven>, 2002.
3. Ralf Nikolai, Ralf Kramer, Marc Steinhaus, Paolo Plini dan Bruno Felluga, "GenThes: A General Thesaurus Browser for Web-based Catalogue Systems", IEEE, Available from: <http://www.computer.org/proceedings/meta/1999/papers/49/rnikolai.html>, 1999.
4. NISO Standard z39.19, "Documentation – Guidelines for the establishment and development of monolingual thesauri", Available from: http://www.chin.gc.ca/English/Standards/vocabulary_thesaurus.html, 1999.