

# Penguraian Bahasa Indonesia dengan Menggunakan Pengurai Collins

Rosa A. Sukamto, Dwi H. Widyantoro  
Sekolah Teknik Elektro dan Informatika  
Institut Teknologi Bandung, Bandung INDONESIA  
rosa\_if\_itb\_01@yahoo.com, dwi@if.itb.ac.id

## Abstract

Tulisan ini berisi hasil penelitian kami menggunakan pengurai Collins untuk menguraikan kalimat berbahasa Indonesia. Kami melakukan adaptasi pada semua *file* masukan pengurai Collins agar dapat digunakan untuk bahasa Indonesia. Pengurai Collins adalah pengurai yang dibuat untuk bahasa Inggris. Kumpulan *file* masukan pengurai Collins antara lain *file* leksikon, *file* grammar, *file events*, *file* simbol non terminal, dan *file* korpus. Kendala terbesar dari penelitian ini adalah tidak adanya *treebank* berbahasa Indonesia. *Treebank* digunakan untuk menghitung probabilitas *grammar*. Untuk menanggulangnya kita dapat melakukan translasi *treebank* berbahasa Inggris ke bahasa Indonesia atau membangun sebuah perkumpulan untuk membuat *treebank* berbahasa Indonesia secara manual.

## 1. Pendahuluan

Pohon pola tata bahasa (*parse tree*) adalah pohon yang merepresentasikan struktur sintaks dari kalimat berdasarkan aturan *grammar*. Pohon pola tata bahasa sangat bermanfaat, misalnya untuk pemeriksaan tata bahasa (*grammar checking*) pada mesin pemrosesan kata (*word processing system*), mesin translasi, mesin penanya dan penjawab (*question answering*), pengekstrak informasi, aplikasi leksikografi, dan pengenalan ucapan (*speech recognizers*).

Pohon pola tata bahasa dapat dibangkitkan secara otomatis dengan menggunakan pengurai (*parser*). Dengan memberikan aturan *grammar* yang benar maka sebuah pengurai (*parser*) akan dapat membangkitkan pohon pola tata bahasa yang benar. Beberapa penelitian telah mengembangkan berbagai algoritma penguraian dengan berbagai pendekatan. Pendekatan yang dilakukan bisa dengan pencarian berbasis hasil (*goal-directed search*) bersifat *top-down* atau

pencarian berbasis data (*data-directed search*) yang bersifat *bottom-up*. Sebuah pengurai *top-down* melakukan pencarian dimulai dari akar hingga daun (kata dalam kalimat). Pengurai *bottom-up* melakukan pencarian dari daun atau kata dalam kalimat ke akar pohon pola tata bahasa.

Pengurai yang sedang berkembang saat ini diantaranya adalah menggunakan pendekatan probabilistik. Pendekatan probabilistik membangun model informasi sintaks dengan menggunakan informasi probabilitas agar pencarian pembangkitan pohon pola tata bahasa lebih efisien. Penguraian dengan menggunakan pendekatan probabilistik cukup efisien untuk mengatasi ambiguitas dengan mengambil nilai probabilitas yang terbesar.

Pengurai Collins [6] merupakan pengurai dengan pendekatan probabilistik yang cukup populer saat ini. Meskipun sangat bermanfaat, tapi semua masukan dan penguraian dikemas untuk bahasa Inggris. Bahasa Indonesia memiliki kesamaan dan perbedaan dengan bahasa Inggris maka sangat memungkinkan menggunakan pengurai Collins untuk bahasa Indonesia.

Tulisan ini terdiri dari beberapa bab. Uraian penelitian-penelitian terkait penguraian dengan metode probabilistik sudah pernah dilakukan pada bab 2. Bab 3 membahas mengenai pengurai Collins. Bab 4 membahas mengenai penyesuaian kumpulan *file* masukan pengurai Collins untuk bahasa Indonesia. Bab 5 berisi diskusi mengenai penelitian kami dan isu-isu terkait. Bab 6 berisi kesimpulan dari penelitian kami.

## 2. Tinjauan Pustaka

Penelitian dalam penguraian dengan pendekatan probabilistik pada tulisan ini diawali dengan penelitian yang dilakukan oleh Schabes dan Water yang mendiskusikan *Stochastic Lexicalized Context-Free*

Grammar (SLCFG) [11] yang juga dikenal sebagai *Probabilistic Lexicalized Context-Free Grammar* (PLCFG) yang merupakan model turunan dari *Probabilistic Context-Free Grammar* (PCFG). Glen Carol mengembangkan SINGER (*Single Reader*) yang menggunakan aturan sebagai masukan dan menggunakan PCFG untuk membangkitkan aturan yang baru. Mark Johnson melakukan penelitian dengan membandingkan PCFG dengan model lain untuk penguraian menggunakan pendekatan probabilistik. Penelitian tersebut menghasilkan bahwa performansi PCFG cukup bagus untuk berbagai kasus.

Charniak [3, 4] membangun sebuah pengurai *bottom-up* untuk bahasa Inggris menggunakan *treebank* (kumpulan pohon kalimat) untuk menghitung probabilitas dari kalimat yang diuraikan. Berikutnya Charniak mengembangkan pengurai *top-down* yang menggunakan *treebank* dan pencarian entropi maksimum [5], mirip dengan menggunakan pohon keputusan. Collins membangun sebuah pengurai dengan pendekatan statistik dengan menghitung kebergantungan kata (*bigram lexical*), berikutnya Collins membangun sebuah pengurai berbasis *head-driven* (pencarian kepala kata pada setiap level pohon pola tata bahasa). Bikel mengembangkan model pengurai berbasis statistik (*framework* pengurai) dengan menggunakan parameter leksikal [2].

Aziz berserta rekan-rekannya mencoba untuk menguraikan bahasa Melayu (Malaysia) menggunakan aturan produksi CFG. Walau kelihatan mirip, bahasa Indonesia dan bahasa Malaysia memiliki perbedaan sehingga hasil dari penelitian tersebut tidak dapat langsung diimplementasikan pada bahasa Indonesia. Lefuel dan Ross mencoba membuat pengurai dengan metode hibrid menggunakan pengurai dengan pendekatan statistik dan algoritma genetik [9]. Jurafsky dan Martin memberikan bahasan yang lebih mendalam mengenai penguraian dengan pendekatan probabilistik untuk bahasa Inggris.

Penelitian yang mirip juga dilakukan oleh Gusmita dan Manurung [8]. Mereka membangun model pengurai dengan menggunakan pendekatan probabilistik untuk bahasa Indonesia dengan menggunakan perangkat PC-PATR untuk membangkitkan pohon pola tata bahasa. Penelitian dalam tulisan ini sama dengan penelitian yang dilakukan Collins, Hajic, Ramshaw dan Tillmann [7]. Mereka mengadaptasi pengurai Collins untuk bahasa Czech sedangkan kami mengadaptasi pengurai Collins untuk bahasa Indonesia.

### 3. Pengurai Collins

Collins memaparkan tiga buah model penguraian dengan pendekatan probabilistik. Pada model yang pertama, PCFG menggunakan aturan produksi berikut:

$$P(h) \rightarrow L_n(l_n) \dots L_1(l_1) H(h) R_1(r_1) \dots R_m(r_m)$$

dimana  $H$  adalah kepala pada aturan  $P$ .  $L_n(l_n) \dots L_1(l_1)$  serta  $R_1(r_1) \dots R_m(r_m)$  adalah aturan sisi kiri dan aturan sisi kanan dilihat dari  $H$ . Collins memberikan parameter jarak (mempertimbangkan posisi simbol non terminal) sebagai pertimbangan pada model ini agar tidak terjadi dominasi sebagian aturan produksi.

Model kedua merupakan lanjutan dari model pertama yang membedakan subkalimat yang berperan sebagai keterangan atau pelengkap (frase keterangan/frase pelengkap). Model ketiga memberikan penanganan terhadap kalimat majemuk dan kalimat tanya. Bagaimana perhitungan probabilitas untuk semua mode serta penjelasan lebih detail dapat dilihat pada disertasi Michael Collins [6].

Pengurai Collins menggunakan PCFG untuk memodelkan pola tata bahasa dan menggunakan algoritma penguraian *chart*. Modul-modul pada pengurai Collins merupakan model-model pembelajaran. Menjalankan pengurai Collins untuk bahasa Indonesia membutuhkan adaptasi dari kumpulan *file* berikut:

1. *File events* untuk menyimpan *events* yang dibangkitkan secara heuristik (probabilitas kebergantungan elemen kalimat) dari WSJ (*Wall Street Journal*) *Treebank* menggunakan format Collins. *File* ini digunakan untuk menghitung probabilitas *grammar* (aturan produksi).
2. Sebuah korpus yang berisi kalimat-kalimat yang telah diberi tag, kalimat-kalimat inilah yang akan dijadikan pohon pola tata bahasa.
3. *File* yang berisi aturan *grammar* sebagai referensi penguraian.
4. *File* yang berisi simbol-simbol non terminal yang digunakan.
5. Sebuah leksikon untuk memeriksa *tag* kata

### 4. Penyesuaian Pengurai Collins untuk Bahasa Indonesia

Bagian ini akan membahas beberapa modifikasi yang diperlukan guna melakukan adaptasi kumpulan

file masukan pengurai Collins untuk bahasa Indonesia sebagai pemrosesan awal (*preprocessing*).

#### 4.1 File Event

Collins membangkitkan *events* dari sebuah bagian di Penn WSJ *treebank* [10]. Bagian ini merupakan bagian yang rumit karena tidak adanya *treebank* berbahasa Indonesia. Hal tersebut akan didiskusikan pada Bab 6. Berikut adalah contoh sebagian dari *file events* berbahasa Indonesia:

```
6 4 Yohanes NN memukul VB Bill NN . PU
3 memukul VB S NP 00000 00000
2 #STOP# #STOP# memukul VB #STOP# S NP
000000 110 0 0
2 #STOP# #STOP# memukul VB #STOP# S NP
000000 010 0 0
3 memukul VB NP VB 00000 00000
2 Yohanes NN memukul VB NN NP VB 000000 110
0 0
2 #STOP# #STOP# memukul VB #STOP# NP VB
000000 100 0 0
2 Bill NN memukul VB NN NP VB 000000 010 0
0
2 . PU memukul VB PU NP VB 000000 000 0 0
2 #STOP# #STOP# memukul VB #STOP# NP VB
000000 000 0 0
```

#### 4.2 File Korpus

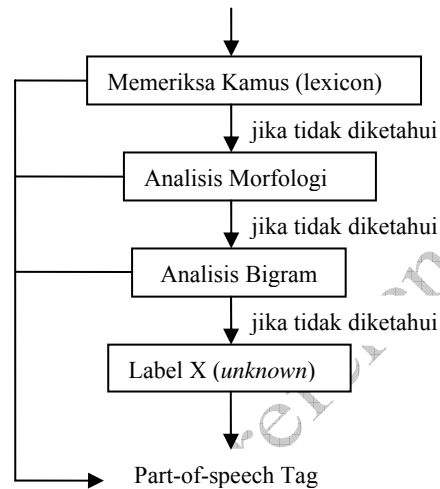
Kalimat yang akan diuraikan perlu diberi tag untuk setiap katanya (*part of speech tagging*). *File* korpus memiliki format sebagai berikut:

```
jumlah_kata kata1 tag1 kata2 tag2 ...
```

seperti contoh berikut:

```
4 Yohanes NN memukul VB Bill NN . PU
      (hits)
```

Gambar 1 merupakan langkah-langkah proses *POS tagging* yang digunakan pada penelitian ini. Pertama kata akan ditentukan tag-nya dengan melihat pada kamus. Jika kata masih belum dapat ditentukan tag-nya maka akan ditentukan dengan analisis morfologi. Jika kata masih belum dapat ditentukan tag-nya maka akan ditentukan dengan analisis bigram menggunakan aturan *grammar* yang ada, misalnya pada frase “sedang menggambar” dimana tag untuk “sedang” adalah RB dan tag untuk “menggambar” tidak diketahui maka jika ada aturan *grammar* VP → RB VB dan setelah dihitung memiliki probabilitas yang terbesar maka dapat disimpulkan bahwa “menggambar” memiliki tag VB.



Gambar 1. POS Tagging

Bahasa Indonesia memiliki aturan morfologi yang dapat digunakan untuk memprediksi kelas kata atau tag kata. Tabel 1 merupakan beberapa aturan morfologi dalam bahasa Indonesia. Urutan proses prediksi tag kata dengan analisis morfologi adalah sebagai berikut:

1. Pemeriksaan kata bilangan; jika kata mengandung angka
2. Pemeriksaan kata singkatan; jika semua huruf pada kata merupakan huruf besar; termasuk kata benda.
3. Pemeriksaan awalan, misalnya “menari” berasal dari kata dasar “tari” dengan awalan *meN* maka akan disimpulkan sebagai kata kerja,
4. Pemeriksaan akhiran, misalnya “terangi” berasal dari kata dasar “terang” mendapat akhiran *-i* yang berarti termasuk kata kerja,
5. Pemeriksaan konfiks (ada awalan dan akhiran),
6. Pemeriksaan pengulangan kata (baik pengulangan kata dasar atau pengulangan berimbuhan), misal “buku-buku” dimana termasuk kata benda (sama dengan jenis kata dasarnya),
7. Pemeriksaan nama; kata yang diawali dengan huruf besar yang berarti termasuk kata benda.

Kami menggunakan 33 aturan prefiks/awalan, 29 aturan sufiks/akhiran, dan 17 konfiks untuk analisis morfologi.

**Tabel 1. Aturan morfologi (imbuan)**

Aturan Morfologi	Contoh Kata	Jenis Kata
meN + kata dasar + kan	mengajukan	kata kerja
peN-ber + kata dasar + an	pelukis, pemburu	kata benda
ke-ber- + kata dasar + an	kebersamaan	kata benda
kata dasar + i	terangi	kata kerja
beR- + kata dasar	bekerja	kata kerja
ter- + kata dasar	tertidur	kata kerja

### 4.3 Grammar dan Simbol Non-Terminal

*File grammar* dibangkitkan dari *treebank*. Seperti halnya *file events*, disini juga ditemukan kendala yaitu tidak adanya *treebank* berbahasa Indonesia. Tata bahasa (*grammar*) pada bahasa Indonesia mirip dengan tata bahasa Inggris seperti adanya subyek-predikat-obyek, tapi tetap saja ada beberapa perbedaan antara bahasa Inggris dan bahasa Indonesia, misalnya kalimat berbahasa Indonesia tidak mengenal perbedaan kata kerja karena waktu kejadian. Bahasa Indonesia memiliki pola frase DM (diterangkan-menerangkan) misalnya buku biru, sedangkan bahasa Inggris memiliki pola frase MD (menerangkan diterangkan) misalnya *blue book*. Kata benda pada bahasa Indonesia juga tidak membedakan benda jamak dan tidak jamak.

Aturan *grammar* pada pengurai Collins yang masih bisa digunakan untuk bahasa Indonesia juga kami gunakan sebagai aturan *grammar* untuk bahasa Indonesia. Contoh tata bahasa (*grammar*) untuk bahasa Indonesia adalah sebagai berikut:

- S → NP VP NN
- NP → NN JJ (misalnya: anak kecil)
- VP → RB VB (misalnya: sedang menulis)

Beberapa simbol non terminal pada pengurai Collins juga perlu dilakukan modifikasi karena tidak semua simbol dapat merepresentasikan bahasa Indonesia. Simbol non terminal yang digunakan pada penelitian ini dapat dilihat pada Tabel 2 dan 3.

Tidak semua tag pada pengurai Collins dapat digunakan untuk bahasa Indonesia. Seperti tag NNP, NNPS, dan NNS yang digunakan pengurai Collins untuk membedakan kata benda jamak dan tidak jamak, sedangkan bahasa Indonesia tidak membedakan kata

benda. Kata benda dalam bahasa Indonesia akan diberi tag NN. Bahasa Indonesia juga tidak mengenal *determiner* seperti kata “the” pada bahasa Inggris. Tag VBD, VBG, VBN, VBP, dan VBZ digunakan pengurai Collins untuk membedakan kata kerja berdasarkan waktu kejadian, sedangkan bahasa Indonesia tidak mengenal kata kerja berdasarkan waktu kejadian. Kata kerja dalam bahasa Indonesia akan diberi tag VB.

**Tabel 2. Simbol non terminal**

Simbol	Jenis Kata	Contoh Kata
JJ	kata sifat	cantik
RB	kata keterangan	nanti, sekarang
AR	artikula	si, sang
CC	konjungsi koordinat	dan, lalu
CS	konjungsi subordinat	ketika, walaupun
PR	pronomina	saya, itu
WH	kata tanya	Siapaakah
NN	kata kerja	meja
CD	kata bilangan	seribu
IN	preposisi	di, ke, dari
UH	kata seru	ai, aah, ceile
RP	partikel	pun, per
VB	kata kerja	melempar
MD	modal	boleh
FW	kata asing	download, notebook
SYM	simbol	+, %, \$, #
PU	tanda baca	., : ; ( ) “ ’ ‘ ’ ’
X	tidak diketahui ( <i>unknown</i> )	

**Tabel 3. Simbol non terminal untuk frase**

Simbol	Description
S	kalimat
ADJP	frase sifat
ADVP	frase keterangan
NP	frase benda
SBAR	sub kalimat
SBARQ	sub kalimat setelah kata tanya
VP	frase kerja

### 4.4 File Leksikon

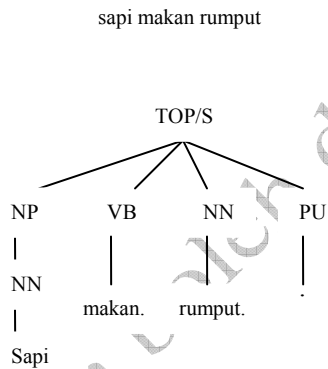
Pengurai Collins menggunakan *file* leksikon untuk memeriksa tag dari kata. Penelitian ini menggunakan KEBI (Kamus Elektronik Bahasa Indonesia) yang dapat digunakan untuk keperluan penelitian. KEBI dikembangkan oleh Badan Pengkajian dan Penerapan Teknologi (BPPT). KEBI berisi 29.396 kata.

KEBI membagi kelas kata/jenis kata mejadi lima belas kelompok yaitu kata sifat (adjektiva), kata

keterangan (adverbia), kata sandang (dibedakan menjadi *determiner* dan *article*), kata bantu (*auxiliary*), kata hubung (konjungsi), kata seru (interjeksi), kata benda (nomina), kata bilangan (dibedakan menjadi *numeral* dan *ordinal*), kata tugas partikel, kata fatis (kata yang menekankan seperti *assalamualaikum*, *bismillah*), kata depan (preposisi), kata ganti (pronomina), dan kata kerja (verba). Kamus tersebut masih harus dikonversikan menjadi format pengurai Collins.

### 5. Eksperimen

Eksperimen yang dilakukan bertujuan untuk memeriksa apakah pengurai Collins dapat digunakan untuk bahasa Indonesia menggunakan kumpulan *file* yang telah diadaptasi untuk bahasa Indonesia. Untuk eksperimen dibuat dua buah kelompok. Kelompok pertama terdiri dari 42 pohon kalimat pada *treebank* dan 7 kalimat sederhana untuk korpus. Pengurai Collins berhasil menguraikan semua kalimat pada kelompok pertama dengan enam kalimat diuraikan dengan benar dan satu kalimat diuraikan dengan kurang tepat. Kurang tepat karena aturan *grammar* yang tepat tidak ada dalam *treebank* yang digunakan. Gambar 2 merupakan salah satu hasil penguraian kalimat dari kelompok pertama.

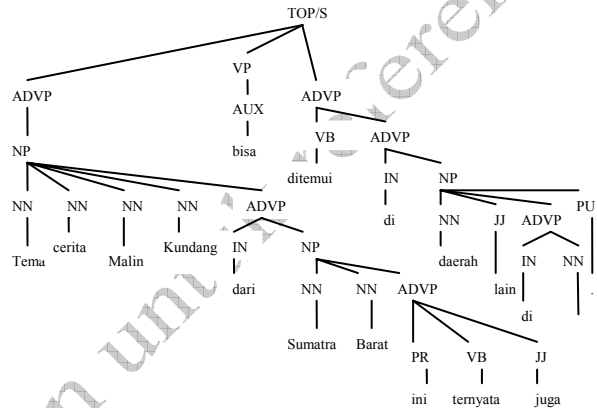


**Gambar 2. Pohon pola tata bahasa dari kalimat sederhana**

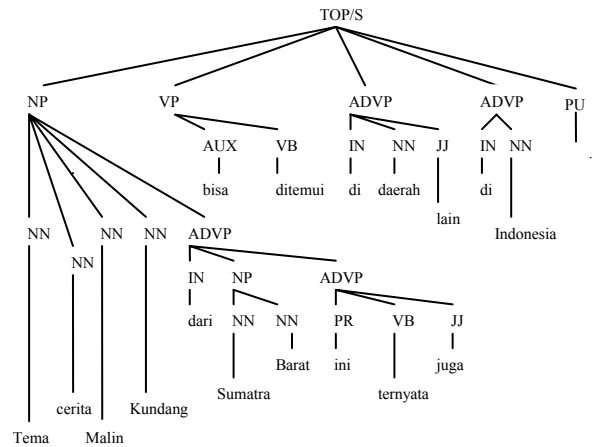
Kelompok kedua terdiri dari 190 pohon kalimat pada *treebank* dan 15 kalimat pada korpus. Pohon kalimat pada *treebank* terdiri dari kalimat sederhana dan kalimat kompleks sedangkan kalimat pada korpus terdiri dari kalimat kompleks. Pengurai Collins dapat menguraikan 8 kalimat dari 15 kalimat. Hasil pohon kalimat pada kelompok kedua yang berhasil diuraikan tidak ada yang benar secara kesatuan kalimat. Kalimat yang tidak dapat diuraikan karena tidak ada *grammar*

yang sesuai dengan bagian kalimat sehingga penguraian tidak dapat menyatukan penguraian kalimat secara keseluruhan. Salah satu contoh kalimat yang dapat diuraikan pengurai Collins pada kelompok kedua dapat dilihat pada Gambar 3, sedangkan hasil yang seharusnya dapat dilihat pada Gambar 4.

Tema cerita Malin Kundang dari Sumatra Barat ini ternyata juga bisa ditemui di daerah lain di Indonesia.



**Gambar 3. Pohon pola tata bahasa dari kalimat kompleks dari pengurai Collins**



**Gambar 4. Pohon pola tata bahasa dari kalimat kompleks yang benar**

## 6. Diskusi

Kendala terbesar dalam penelitian ini adalah *treebank* yang digunakan untuk membangkitkan *file events* untuk menghitung probabilitas aturan *grammar*. Sayangnya belum ada *treebank* untuk bahasa Indonesia. Penelitian ini mencoba membuat *treebank* dengan jumlah yang sangat kecil jika dibandingkan dengan Penn *treebank*. Merupakan tantangan terbesar bagi bangsa ini untuk membuat sebuah *treebank* berbahasa Indonesia yang dapat mendukung penelitian-penelitian mengenai bahasa Indonesia. Ada dua cara yang dapat dilakukan untuk membuat *treebank* berbahasa Indonesia.

Cara pertama adalah melakukan translasi terhadap Penn *treebank* ke bahasa Indonesia. Cara ini bisa jadi menjadi cara yang paling mudah dan efisien untuk membuat Penn *treebank* berbahasa Indonesia, tapi ketidakkurasian tata bahasa masih perlu diperiksa kembali agar sesuai dengan aturan bahasa Indonesia. Penanganan khusus juga perlu dilakukan untuk menangani kata berbahasa Inggris yang tidak ada dalam bahasa Indonesia (misalnya diganti dengan *blank*). Misal sebagai contoh salah satu contoh kalimat pada Penn *treebank* berikut:

(S (NP-SBJ (NNP President)  
(NNP Reagan))  
(VP (VBD learned)  
(NP (DT that)  
(NN lesson)))  
(. .))

ditranslasikan menjadi bahasa Indonesia, maka akan menjadi sebagai berikut:

(S (NP (NN Presiden)  
(NN Reagan))  
(VP (VB belajar)  
(NP (NN pelajaran)  
(PR itu)))  
(. .))

Cara yang kedua adalah dengan membuat sebuah permainan yang dapat diakses semua masyarakat Indonesia, misalnya permainan “Apakah Anda Benar-benar Orang Indonesia” dimana di dalamnya berisi pertanyaan untuk membuat pohon kalimat dengan batasan waktu. Waktu yang tersisa akan dijadikan skor permainan. Orang-orang biasanya lebih tertarik pada permainan dibandingkan sesuatu yang serius untuk membuat *treebank*.

Membuat *treebank* secara manual merupakan pekerjaan yang tidak mudah dan tidak dapat dilakukan secara individu. *Treebank* sangat bermanfaat untuk kemajuan penelitian bangsa ini. Sebagai langkah awal, kami merencanakan untuk membuat sebuah sistem berbasis web untuk mengelola dan memelihara pembuatan *treebank*. Langkah berikutnya adalah bagaimana meningkatkan motivasi masyarakat untuk ikut berpartisipasi pada pembuatan *treebank* berbahasa Indonesia.

## 7. Kesimpulan

Secara teori sangat memungkinkan menggunakan pengurai Collins untuk bahasa Indonesia karena semua *file* masukan dapat diadaptasi ke bahasa Indonesia. Kendala terbesar dari penelitian ini adalah tidak adanya *treebank* untuk bahasa Indonesia sehingga hasil eksperimen awal kurang konsisten. Oleh karena itu sangat perlu dibuat *treebank* berbahasa Indonesia beserta korpus berbahasa Indonesia untuk memajukan penelitian di bidang bahasa Indonesia.

## 8. Referensi

- [1] Azis, Mohd Juzaidin et al. (2006) Pola Grammar Technique for Grammatical Relation Extraction of Malay Language, Malaysian Journal of Computer Science, 19, 59-72
- [2] Bikel, Daniel M. (2004) : *On The Parameter Space of Generative Lexicalized Statistical Parsing Models*, Disertasi, University of Pennsylvania. 1-20, 141-148
- [3] Charniak, Eugene. (1993) : *Statistical Language Learning*, Massachusetts Institute of Technology.
- [4] Charniak, Eugene. (1997) : *Statistical Parsing with a Context-free Grammar and Word Statistics*, American Association for Artificial Intelligence: AAAI Press. 1-6
- [5] Charniak, Eugene. (2000) : *A Maximum-Entropy-Inspired Parser*, *Proceedings of NAACL-2000*. 132-139.
- [6] Collins, Michael. (1999) : *Head-Driven Statistical Models for Natural Language Parsing*, Disertasi program Doctor of Philosophy, University of Pennsylvania. 1-265.
- [7] Collins, Michael, Jan Hajic, Lance Ramshaw, Cristoph Tillmann (1999) : *A Statistical Parser for Czech*, *Proceedings of the 37th Annual Meeting of the ACL*.

- [8] Gusmita, Ria Hari & Ruli Manurung (2008) Some initial experiments with Indonesian probabilistic parsing. Second MALINDO Workshop. 1-5.
- [9] Lefuel, Ramon & Brian J. Ross (2004) Parsing Probabilistic Context Free Language with Multiple-Objective Genetic Algorithm, Technical Report, Brock University. 1-11.
- [10] Marcus, Mitchell P. dkk (1992) : Building a Large Annotated Corpus of English: The Penn Treebank. Departmet of Computer and Information Science University of Pennsylvania.
- [11] Schabes, Yves & Waters, Richard C (1993) Stochastic Lexicalized Context-Free Grammar, International Workshop on Parsing Technology. 1-10.

Hanya boleh digunakan untuk referensi