



melalui proses tokenisasi. Kemudian, hasil tokenisasi digunakan untuk identifikasi bagian-bagian UU, yaitu bagian judul, pembukaan, batang tubuh, dan penutup. Dalam proses identifikasi judul, juga dilakukan proses identifikasi ID dokumen. Hal ini dapat dilakukan karena pada bagian judul terdapat nomor dan tahun UU yang digunakan untuk membentuk ID dokumen.

Proses selanjutnya adalah pemrosesan bagian judul. Pada proses ini dilakukan identifikasi subbagian dari judul untuk kemudian dibuat elemen-elemen XML yang bersesuaian. Proses ini juga memanggil proses pembuatan referensi untuk mengecek apakah ada pengacuan ke dokumen legal lain atau tidak. Jika ditemukan adanya frase pengacuan maka akan dibuat elemen XML-nya dan ditempatkan kembali di lokasi ditemukannya frase pengacuan tersebut.

Pemrosesan bagian pembukaan dilakukan setelah bagian judul selesai diproses. Pemrosesannya meliputi identifikasi subbagian beserta ID-nya dan kemudian dibuat elemen XML-nya. Bagian pembukaan dapat mengandung frase pengacuan atau referensi ke bagian lain atau ke dokumen legal lain. Untuk itu, perlu dilakukan pemanggilan proses pembuatan referensi. Proses pembuatan referensi akan menghasilkan frase-frase pengacuan yang selanjutnya diproses untuk membuat elemen-elemen XML yang menandakan adanya keterkaitan. Elemen-elemen pengacuan tersebut kemudian digabung dengan elemen-elemen lain sesuai lokasi penemuan frase pengacuannya.

Pemrosesan bagian batang tubuh dilakukan berdasarkan hasil dari proses identifikasi jenis UU. Proses identifikasi jenis UU dilakukan berdasarkan elemen XML yang terdapat pada bagian judul. Jika UU yang diproses termasuk UU non-perubahan maka dipanggil proses bagian batang tubuh UU non-perubahan. Sedangkan, jika UU yang diproses termasuk UU perubahan maka dipanggil proses bagian batang tubuh UU perubahan.

Pemrosesan bagian batang tubuh, baik batang tubuh UU non-perubahan maupun batang tubuh UU non-perubahan, meliputi identifikasi subbagian beserta ID-nya dan kemudian dibuat elemen XML-nya. Pada bagian ini juga dapat ditemukan frase pengacuan atau referensi ke bagian lain atau ke dokumen legal lain sehingga perlu dilakukan pemanggilan proses pembuatan referensi. Proses pembuatan referensi akan menghasilkan frase-frase pengacuan yang

selanjutnya diproses untuk membuat elemen-elemen XML yang menandakan adanya keterkaitan. Elemen-elemen pengacuan tersebut kemudian digabung dengan elemen-elemen lain sesuai lokasi penemuan frase pengacuannya.

Proses selanjutnya adalah proses bagian penutup. Pemrosesan bagian ini sama saja untuk UU non-perubahan dan UU perubahan yang meliputi identifikasi subbagian beserta ID nya dan pembuatan elemen XML dari subbagian tersebut. Setelah seluruh bagian UU diproses maka akan dihasilkan empat elemen induk utama, yaitu JUDUL, PEMBUKAAN, BATANG\_TUBUH, dan PENUTUP dengan sejumlah elemen anak di dalamnya. Elemen-elemen tersebut kemudian disatukan untuk membentuk dokumen UU dalam format XML. Dalam penyatuannya, elemen DOKUMEN digunakan sebagai elemen akar.

Dokumen UU dalam format XML digunakan sebagai dokumen masukan pada fitur pembuatan graf referensi. Sama seperti halnya fitur standarisasi dokumen, pada fitur ini dokumen masukan juga dibaca dan diuraikan menjadi token-token. Selanjutnya dilakukan pencarian elemen XML yang menandakan adanya pengacuan atau referensi. Kemudian dilakukan pembuatan graf dengan elemen-elemen XML yang ditemukan bertindak sebagai *node* dari graf dan hubungannya dengan elemen lain ditandai dengan *edge*. Properti graf yang sudah dideklarasikan tersebut disimpan dalam dokumen berformat DOT sebagai keluaran dari fitur pembuaran graf referensi ini.

Subbab – subbab berikutnya menjelaskan setiap bagian dari arsitektur sistem yang ada pada Gambar 3.1 di atas.

### **3.2 Dokumen Masukan dan Dokumen Keluaran**

Fitur pertama dari sistem yaitu fitur standarisasi dokumen menerima masukan berupa dokumen UU dalam format teks. Dokumen UU ini dikelompokkan menjadi dua, yakni dokumen UU perubahan dan dokumen UU non-perubahan. Contoh dari masing-masing dokumen tersebut dapat dilihat pada Lampiran 4 dan 5. Agar dapat diproses, dokumen masukan setidaknya harus mengikuti format sistematika penyusunan UU.

Keluaran dari fitur standarisasi ini adalah dokumen UU dalam format XML, yaitu dokumen yang terdiri dari sejumlah elemen XML yang

merepresentasikan bagian-bagian beserta subbagian dari UU. Contoh dokumen keluaran dapat dilihat pada Lampiran 6 dan 7.

Dokumen keluaran dari fitur standarisasi digunakan sebagai masukan bagi fitur pembuatan graf referensi dokumen. Dokumen UU dalam format XML tersebut diproses untuk menghasilkan dokumen dalam format DOT yang berisi deskripsi graf referensi. Contoh dokumen DOT yang dihasilkan dapat dilihat pada Lampiran 8.

### 3.3 Kamus

Fitur standarisasi dokumen menggunakan kamus yang memuat informasi-informasi dibutuhkan selama pemrosesan dokumen masukan. Kamus tersebut terbagi menjadi tiga, yaitu kamus kata hubung, kamus jabatan, dan kamus nama pejabat. Kamus kata hubung diperlukan pada pemrosesan bagian batang tubuh. Kamus ini menyimpan kata-kata hubung yang digunakan untuk menghubungkan kata, frase, atau kalimat menjadi satu kalimat yang utuh. Contohnya, kata 'dan', 'atau', dan 'kecuali'.

Sedangkan kamus jabatan dan nama pejabat diperlukan pada pemrosesan bagian penutup. Kamus jabatan berisi nama-nama jabatan pemerintahan Indonesia yang berkaitan dengan pembuatan UU, seperti 'Presiden Republik Indonesia' dan 'Sekretaris Negara Republik Indonesia'. Kamus nama pejabat berisi nama-nama pejabat pemerintah yang bertugas untuk mengesahkan/menetapkan dan mengundang UU. Contohnya, 'Susilo Bambang Yudhoyono' sebagai Presiden Republik Indonesia periode 2004-2009 dan 'Andi Mattalatta' selaku Menteri Hukum dan Hak Asasi Manusia Republik Indonesia pada periode yang sama.

Ketiga kamus tersebut di atas disimpan dalam format teks. Kamus dibuat secara manual berdasarkan analisis terhadap dokumen observasi. Dalam penelitian ini digunakan 60 dokumen observasi yang terdiri dari 40 dokumen UU non-perubahan dan 20 dokumen UU perubahan. Karakteristik dari dokumen observasi yang digunakan adalah sebagai berikut:

- UU pada rentang tahun 1998 sampai tahun 2008
- Ukuran berkas berkisar antara 4 KB sampai 99 KB

- Jumlah pasal pada bagian batang tubuh UU non-perubahan berkisar antara 2 sampai 93 pasal
- Jumlah ketentuan yang diubah pada pada pasal I UU perubahan berkisar antara 1 sampai 59 ketentuan
- Jumlah frase pengacuan berkisar antara 13 frase sampai 194 frase

Analisis bagian batang tubuh untuk membuat kamus kata hubung dilakukan dengan cara mencari kata hubung yang melekat pada kata ‘bab’, ‘bagian’, ‘paragraf’, ‘pasal’, ‘ayat’, ‘huruf’, dan ‘angka’. Kata-kata hubung tersebut nantinya dapat membantu dalam menemukan subbagian dari batang tubuh. Sedangkan analisis bagian penutup untuk membuat kamus jabatan dan nama pejabat dilakukan pada bagian pengesahan/penetapan dan bagian pengundangan.

### 3.4 Aturan

Aturan digunakan untuk memproses setiap bagian dokumen UU. Aturan-aturan tersebut dibuat dengan cara menganalisis dokumen observasi. Proses analisis dilakukan dengan cara mencari pola-pola untuk mengenali objek-objek yang ada dalam dokumen. Objek yang dimaksud adalah subbagian yang dimiliki setiap bagian UU serta frase-frase yang menunjukkan adanya pengacuan.

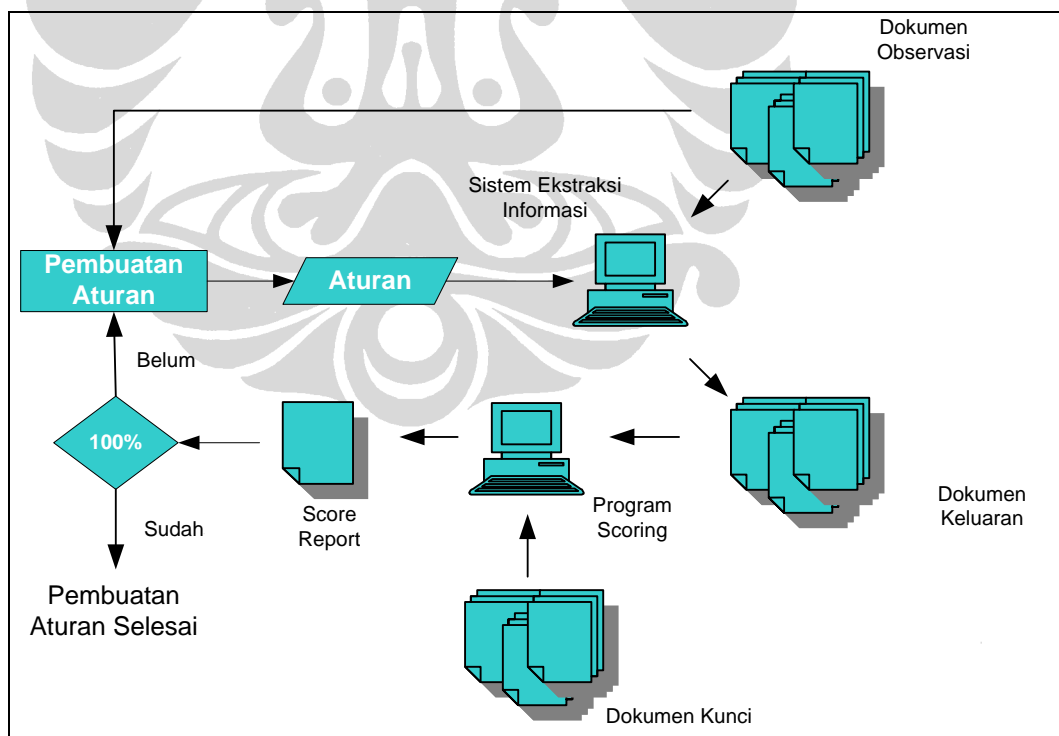
Proses pembuatan aturan dibantu oleh informasi kontekstual yang terdapat dalam dokumen. Informasi tersebut meliputi informasi kata kunci, informasi kata hubung, informasi tanda baca, dan informasi frase khusus. Informasi kata kunci berupa kata-kata yang menunjukkan secara langsung bagian dan subbagian dari dokumen. Contohnya, kata kunci ‘bab’ menunjukkan subbagian bab dan kata kunci ‘mengingat’ menunjukkan subbagian dasar hukum. Kata-kata kunci tersebut akan dijelaskan lebih lanjut pada subbab berikutnya.

Informasi kata hubung adalah kata hubung yang melekat atau berdampingan dengan kata kunci. Kata-kata hubung ini disimpan dalam kamus tersendiri yaitu kamus kata hubung seperti yang telah dijelaskan di atas. Selain itu, terdapat juga informasi tanda baca yang dapat memberikan arti khusus pada suatu kalimat atau kata. Tanda baca yang diperhatikan dalam penelitian ini adalah tanda baca titik (‘.’), koma (‘,’), titik dua (‘:’), titik koma (‘;’), dan kurung (‘(’ dan ‘)’).

Contoh penggunaan tanda baca adalah dalam penentuan subbagian ayat, yaitu dengan mencari angka Arab (1, 2, 3, dan seterusnya) yang diapit tanda baca kurung.

Informasi terakhir yang dibutuhkan adalah informasi frase khusus. Informasi ini berupa gabungan beberapa kata yang membentuk suatu arti tertentu. Contohnya adalah frase nama jabatan dan frase nama pejabat yang disimpan dalam kamus. Contoh lain seperti frase 'Dengan Rahmat Tuhan Yang Maha Esa' yang menunjukkan subbagian frase syukur pada bagian pembukaan.

Proses pembuatan aturan dilakukan secara iteratif. Artinya, aturan yang sudah dibuat diimplementasikan pada sistem untuk kemudian diuji pada dokumen observasi. Hasilnya dianalisis untuk mengetahui apakah aturan yang dibuat sudah mencukupi atau belum dengan cara membandingkan dokumen keluaran sistem dengan dokumen yang dibuat secara manual. Jika hasil perbandingan yang terdiri dari nilai *recall*, *precision*, dan *F-measure* belum mencapai 100%, maka pembuatan aturan dilanjutkan sampai ketiga nilai tersebut mencapai nilai sempurna. Proses pembuatan aturan dapat dilihat pada gambar di bawah ini:



Gambar 3.2: Proses Pembuatan Aturan

### 3.5 Tokenisasi

Proses pertama dari fitur standardisasi dokumen adalah tokenisasi. Tokenisasi adalah proses pembacaan dokumen masukan untuk kemudian diuraikan menjadi token-token. Token yang dihasilkan berupa kata-kata tanpa spasi beserta karakter-karakter legal XML. Karakter-karakter legal XML yang dimaksud adalah karakter-karakter yang tidak menyebabkan *error* jika dieksekusi di *browser*. Karakter yang dapat menyebabkan *error* seperti '<', '?', dan '&'.

Pada proses ini juga dilakukan penanganan terhadap token 'undang-undang' karena mengingat dokumen masukan yang tidak selalu mengikuti penulisan bahasa Indonesia yang baik dan benar. Adakalanya kata 'undang-undang' ditulis 'undangundang', 'undang undang', 'undang- undang', dan 'undang -undang'. Jadi untuk menyamakan, semuanya diubah menjadi kata 'undang-undang'. Hal ini diperlukan agar proses-proses selanjutnya bisa berjalan dengan benar. Misalnya pada proses identifikasi frase pengacuan, jika penulisan kata 'undang-undang' tidak benar, maka frase tidak akan teridentifikasi.

### 3.6 Identifikasi Bagian UU

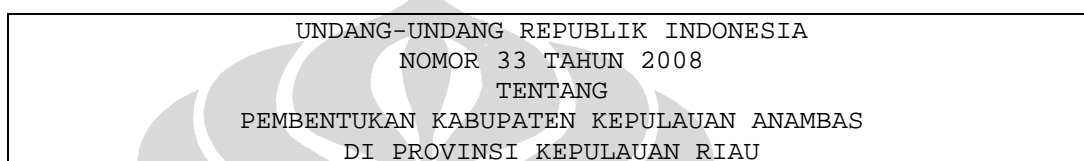
Setelah proses tokenisasi dilakukan, proses selanjutnya adalah identifikasi bagian UU. Seperti yang telah dijelaskan sebelumnya, bahwa struktur UU terdiri dari bagian judul, pembukaan, batang tubuh, dan penutup. Untuk mengidentifikasi setiap bagian tersebut, ditentukan token penanda batas akhir dari masing-masingnya. Token penanda batas akhir suatu bagian ditentukan berdasarkan token penanda batas awal dari bagian selanjutnya. Contohnya, jika ingin mencari token penanda batas akhir bagian judul maka dicari token penanda batas awal dari bagian pembukaan.

Proses identifikasi bagian UU pada penelitian ini sama dengan proses identifikasi bagian UU yang dilakukan Mawaddah (2006). Penjelasan proses ini selengkapnya dapat dilihat di Mawaddah (2006, p 60-61).

Pada proses identifikasi bagian judul juga dilakukan identifikasi ID dokumen. ID dokumen terdiri dari dua komponen, yaitu nomor UU dan tahun pengesahan/penetapan UU yang keduanya terdapat pada bagian judul. ID

dokumen berfungsi sebagai pembeda antara satu dokumen dengan dokumen lain. Selain itu, ID ini juga dibutuhkan pada proses-proses selanjutnya.

Format ID dokumen yang digunakan pada penelitian ini adalah gabungan dari singkatan dari undang-undang, nomor undang-undang, dan tahun pengesahan/penetapan undang-undang yang dipisahkan dengan tanda baca '-'. Jadi, format ID-nya adalah 'uu-[nomor UU]-[tahun UU]'. Nomor UU diperoleh dari angka di belakang kata 'NOMOR' dan tahun UU diperoleh dari angka di belakang kata 'TAHUN'. Misalnya, akan dibentuk ID dokumen dari UU yang bagian judulnya seperti terlihat pada gambar di bawah ini:



**Gambar 3.3: Contoh Bagian Judul UU**

Berdasarkan bagian judul UU yang terdapat pada Gambar 3.3, maka akan diperoleh ID dokumen 'uu-33-2008'.

### 3.7 Proses Bagian Judul

Pemrosesan bagian judul dilakukan berdasarkan token-token yang telah diperoleh pada proses identifikasi bagian UU. Proses ini terdiri dari dua sub proses, yaitu sub proses identifikasi subbagian dan sub proses pembuatan elemen XML. Karena pada bagian judul dapat ditemukan keterkaitan/referensi ke peraturan perundang-undangan lain, maka pada sub proses pembuatan elemen XML juga dilakukan pemanggilan proses pembuatan referensi.

Proses pembuatan referensi bertugas mencari frase pengacuan yang menunjukkan adanya keterkaitan. Keterkaitan pada bagian judul biasanya ditemukan pada UU jenis UU perubahan, UU penetapan peraturan pemerintah pengganti UU, UU pencabutan UU, dan UU pencabutan peraturan pemerintah pengganti UU. Frase-frase yang ditemukan kemudian diubah menjadi elemen-elemen XML yang mengandung atribut ID dari peraturan perundang-undangan yang menjadi target referensi.

Berdasarkan analisis yang telah dilakukan, terdapat tiga subbagian pada bagian judul, yaitu subbagian nomor, tahun, dan nama peraturan. Penjelasan dari



karakteristik setiap bagian tersebut beserta aturan identifikasinya selengkapnya dapat dilihat pada Mawaddah (2006, p 63-65).

Setelah seluruh subbagian dari judul teridentifikasi, proses selanjutnya adalah pembuatan elemen XML setiap subbagian tersebut. Proses pembuatan elemen XML bagian judul pada penelitian ini agak berbeda dengan pembuatan elemen XML bagian judul yang dilakukan Mawaddah (2006). Pada penelitian ini, pembuatan elemen XML dari subbagian nama peraturan memanggil proses pembuatan referensi, sedangkan Mawaddah tidak melakukan hal tersebut.

Elemen XML dibuat dengan cara memberikan *tag* pembuka, isi elemen, dan *tag* penutup. *Tag* pembuka diberikan dengan mengubah format *tag* penanda menjadi *tag* bernama sesuai tipe subbagian. Pembuatan elemen XML bagian judul mengikuti aturan seperti yang terlihat pada Tabel 3.1.

**Tabel 3.1:** Aturan Pembuatan Elemen XML Bagian Judul

Aturan	
<b>1. Subbagian NOMOR</b>	
JIKA	Token[i]="<SUB_BAGIAN TYPE="NOMOR">"
MAKA	IsiElemen[i]=Token[i+1] Elemen[i]="<NOMOR>" + IsiElemen[i] + "</NOMOR>"
<b>2. Subbagian TAHUN</b>	
JIKA	Token[i]="<SUB_BAGIAN TYPE="TAHUN">"
MAKA	IsiElemen[i]=Token[i+1] Elemen[i]="<TAHUN>" + IsiElemen[i] + "</TAHUN>"
<b>3. Subbagian NAMA PERATURAN</b>	
JIKA	Token[i]="<SUB_BAGIAN TYPE="NAMA PERATURAN">"
MAKA	IsiElemen[i]=Gabungan_Semua_Token_Setelah_Token[i]_Sampai_Token_Batas_Akhir IsiElemen[i]=PEMBUATAN_REFERENSI(IsiElemen[i], "") Elemen[i]="<NAMA PERATURAN>" + IsiElemen[i] + "</ NAMA PERATURAN >"

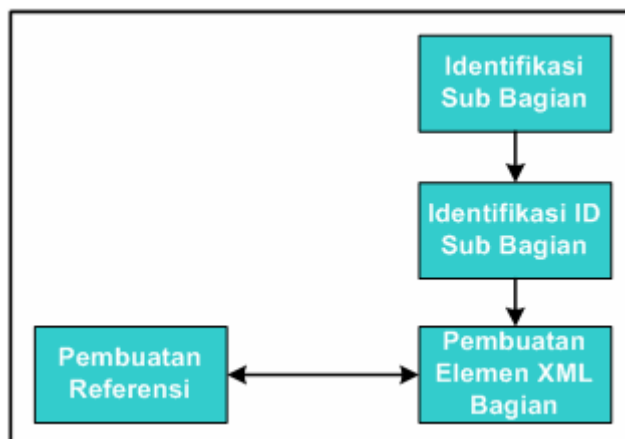
Keterangan: Bagian yang dicetak hitam tebal adalah bagian yang berbeda dengan penelitian yang dilakukan Mawaddah (2006)

Elemen XML dari subbagian nomor dibuat dengan cara mengganti *tag* penandanya dengan *tag* pembuka <NOMOR>, ditambah isi elemen, serta *tag* penutup </NOMOR>. Isi dari elemen ini adalah token yang berada tepat setelah token yang berisi *tag* penanda. Begitu juga dengan elemen XML dari subbagian tahun. *Tag* penandanya diganti dengan *tag* pembuka <TAHUN> ditambah isi elemen berupa token yang berada tepat setelah *tag* penanda, dan diikuti *tag* penutup </TAHUN>.

Pada pembentukan elemen XML dari subbagian nama peraturan dilakukan pemanggilan aturan `PEMBUATAN_REFERENSI` yang terdapat dalam proses pembuatan referensi. Hal ini diperlukan karena mengingat di subbagian ini dapat ditemukan keterkaitan UU dengan peraturan perundang-undangan lain. Sejauh ini keterkaitan yang dapat ditangani adalah keterkaitan suatu UU dengan UU lain dan Kitab Undang-Undang Pidana (KUHP). Proses pembuatan referensi akan menerima masukan berupa isi elemen dari subbagian nama peraturan, yaitu berupa gabungan dari semua token yang terdapat setelah *tag* penandanya sampai token terakhir dari bagian judul. Hasil dari proses tersebut akan menjadi isi sebenarnya dari elemen subbagian nama peraturan yang diapit *tag* pembuka `<NAMA_PERATURAN>` dan *tag* penutup `</NAMA_PERATURAN>`.

### 3.8 Proses Bagian Pembukaan

Sama seperti bagian judul, pemrosesan bagian pembukaan dilakukan berdasarkan token-token yang telah diperoleh pada proses identifikasi bagian UU. Pemrosesan bagian ini terdiri dari dua sub proses, yaitu sub proses identifikasi subbagian dan sub proses pembuatan elemen XML. Sub proses pembuatan elemen XML memanggil proses pembuatan referensi, karena pada bagian ini dapat ditemukan keterkaitan UU dengan peraturan perundang-undangan lain atau keterkaitan antar-subbagian dalam UU itu sendiri. Untuk mendukung proses pembuatan referensi, pada bagian pembukaan dilakukan sub proses identifikasi dan pembuatan ID subbagian dari masing-masing bagian. Alur dari sub proses tersebut dapat dilihat pada Gambar 3.4. Alur tersebut juga berlaku dalam pemrosesan bagian batang tubuh, baik batang tubuh UU non-perubahan maupun batang tubuh UU perubahan.



**Gambar 3.4: Alur Dari Proses Bagian Pembukaan dan Batang Tubuh**

Sumber: Mawaddah, S. (2006). Pengembangan standar dokumen legal indonesia berbasis XML menggunakan sistem ekstraksi informasi dengan studi kasus dokumen undang-undang republik Indonesia.

Untuk dapat mengidentifikasi subbagian dalam bagian pembukaan diperlukan analisis terhadap karakteristik yang dimilikinya. Karakteristik tersebut meliputi posisi subbagian tersebut dalam bagian serta informasi kontekstual yang dapat membantu mengidentifikasi subbagian tersebut. Berdasarkan analisis yang telah dilakukan, ditemukan sembilan subbagian pada bagian pembukaan, yaitu subbagian frase syukur, pejabat pembuat, konsiderans, huruf, dasar hukum, angka, persetujuan, memutuskan, dan menetapkan. Penjelasan karakteristik dari setiap subbagian tersebut selengkapnya dapat dilihat pada Mawaddah (2006, p 65-66).

Berdasarkan karakteristik yang dimiliki setiap bagian, maka dapat dilakukan identifikasi subbagian dari pembukaan. Token penanda setiap bagian diganti dengan *tag* penanda yang sesuai berdasarkan aturan-aturan yang telah dibuat oleh Mawaddah (2006). Penjelasan dari setiap aturan tersebut secara rinci dapat dilihat pada Mawaddah (2006, p 129-131).

Setelah semua subbagian teridentifikasi, proses selanjutnya adalah identifikasi ID setiap subbagian tersebut. Pemberian ID diberikan pada subbagian konsiderans, dasar hukum, angka, dan huruf karena hanya pada subbagian tersebut terdapat frase pengacuan atau menjadi target pengacuan. ID yang diberikan pada suatu subbagian memiliki format berupa gabungan antara ID dokumen dengan nama subbagian bersangkutan, misalnya subbagian konsiderans dan dasar hukum. Apabila subbagian tersebut berada dalam subbagian lain sebagai induknya, maka format ID nya adalah berupa gabungan ID subbagian induk dengan ID subbagian

itu sendiri. Contohnya, subbagian ANGKA yang berada dalam subbagian DASAR\_HUKUM memiliki ID berupa gabungan ID subbagian dasar hukum dan ID subbagian angka itu sendiri. Berikut daftar format ID setiap subbagian pada bagian pembukaan:

**Tabel 3.2: Daftar Format ID Subbagian pada Bagian Pembukaan**

SUBBAGIAN	SUBBAGIAN INDUK	FORMAT ID
KONSIDERANS	-	"[ID_dokumen].konsiderans"
HURUF	KONSIDERANS	"[ID_konsiderans].hr-[no_huruf]"
DASAR_HUKUM	-	"[ID_dokumen].dasar_hukum"
ANGKA	DASAR_HUKUM	"[ID_dasar_hukum].ag-[no_angka]"

Pada proses identifikasi ID subbagian juga dilakukan pemberian atribut NO yang merepresentasikan nomor urut subbagian. Atribut NO ini diberikan pada subbagian angka dan huruf. Isi dari atribut NO diperoleh dari nomor urut subbagian yang melekat pada *tag* penanda atau dari token yang berada setelah *tag* penanda. Aturan dalam proses identifikasi ID dan pemberian atribut NO pada subbagian yang berada dalam bagian pembukaan mengikuti aturan yang telah dikembangkan oleh Mawaddah (2006). Penjelasan dari setiap aturan tersebut dapat dilihat selengkap pada Mawaddah (2006, 137).

Proses terakhir pada bagian pembukaan adalah pembuatan elemen XML. Pada proses ini dilakukan pembuatan elemen-elemen XML sesuai dengan subbagian yang ditemukan. Elemen XML terdiri dari *tag* pembuka, isi elemen, dan *tag* penutup. *Tag* pembuka subbagian dibuat dengan cara mengubah format *tag* penanda subbagian tersebut menjadi *tag* bernama. Contohnya, *tag* penanda subbagian konsiderans, yaitu <SUB\_BAGIAN TYPE="FRASE\_SYUKUR"> diubah menjadi *tag* bernama <FRASE\_SYUKUR>.

Isi elemen dibuat dengan memperhatikan karakteristik setiap subbagian yang berada dalam bagian. Jika suatu subbagian tidak memiliki subbagian anak di dalamnya maka isi elemen XML nya adalah berupa gabungan token-token yang berada setelah *tag* penandanya sampai token sebelum *tag* penanda berikutnya. Contohnya isi elemen frase syukur adalah gabungan token-token setelah *tag* penanda subbagian frase syukur sampai token sebelum *tag* penanda subbagian pejabat pembuat.

Seperti yang telah dijelaskan sebelumnya bahwa pada bagian pembukaan dapat ditemukan frase-frase pengacuan. Frase pengacuan ditemukan pada subbagian huruf dan angka. Oleh karena itu, isi elemen subbagian tersebut harus diproses kembali melalui proses pembuatan referensi. Untuk keperluan proses tersebut, di samping mengirim isi elemen, proses pembuatan elemen XML juga mengirim ID subbagian beserta ID subbagian induknya. Jadi, proses pembuatan referensi akan menerima input berupa ID subbagian huruf dan ID subbagian konsiderans serta ID subbagian angka dan ID subbagian dasar hukum.

Isi elemen XML yang telah selesai diproses oleh proses pembuatan referensi dikirimkan kembali ke proses pembuatan elemen XML. Selanjutnya elemen-elemen subbagian tersebut diberi *tag* penutup yang bersesuaian dengan *tag* pembukanya. Contohnya, elemen XML subbagian frase syukur dengan *tag* pembuka <FRASE\_SYUKUR> diberi *tag* penutup </FRASE\_SYUKUR>. Selanjutnya, elemen-elemen pada bagian pembukaan diberi elemen induk dengan cara meletakkannya di antara *tag* pembuka <PEMBUKAAN> dan *tag* penutup </PEMBUKAAN>. Aturan pembuatan elemen XML pada penelitian ini mengacu pada aturan yang dikembangkan Mawaddah (2006). Penjelasan selengkapnya mengenai aturan-aturan tersebut dapat dilihat pada Mawaddah (2006, p 141-145).

### 3.9 Identifikasi Jenis UU

Seperti yang telah dijelaskan pada bab sebelumnya, UU dapat dibedakan menjadi dua jenis berdasarkan sistematika dan format penulisannya, yaitu UU perubahan dan UU non-perubahan. Suatu UU dapat diketahui termasuk ke jenis yang mana dari bagian judul UU tersebut. Jadi, hasil dari pemrosesan bagian judul dapat membantu proses identifikasi jenis UU.

Proses identifikasi jenis UU diperlukan untuk pemrosesan bagian batang tubuh. Hal ini dikarena sistematika dan format penulisan yang berbeda di antara kedua jenis UU pada bagian tersebut, sehingga diperlukan pemrosesan yang berbeda pula.

Jenis suatu UU dapat diketahui dari elemen XML subbagian nama peraturan. Isi elemen XML pada subbagian ini dapat memuat frase di bawah ini:

- PERUBAHAN ATAS <REF ID=
- PERUBAHAN BERLAKUNYA <REF ID=

- PERUBAHAN <REF ID=
- PERUBAHAN KEDUA ATAS <REF ID=
- PERUBAHAN KETIGA ATAS <REF ID=

Jika isi elemen XML subbagian nama peraturan suatu UU mengandung salah satu frase di atas, maka UU tersebut tergolong UU perubahan. Namun, jika isi elemen XML subbagian nama peraturan tidak memuat salah satu dari frase tersebut maka UU tergolong UU non-perubahan. Jika jenis dokumen UU yang diproses adalah UU non-perubahan proses yang dipanggil adalah proses bagian batang tubuh UU non-perubahan. Sedangkan, jika jenis dokumen UU yang diproses adalah UU perubahan maka proses selanjutnya yang dipanggil adalah proses bagian batang tubuh UU perubahan.

### **3.10 Proses Bagian Batang Tubuh UU Non-Perubahan**

Hasil analisis menunjukkan bahwa terdapat tujuh subbagian pada batang tubuh UU non-perubahan, yaitu subbagian bab, bagian, paragraf, pasal, ayat, huruf, dan angka. Penjelasan karakteristik dari setiap subbagian tersebut dapat dilihat pada Mawaddah (2006, p 67-68). Dari masing-masing karakteristik yang dimiliki setiap subbagian, maka dapat dibuat aturan-aturan untuk mengidentifikasi subbagian tersebut. Aturan tersebut dapat dilihat pada Mawaddah (2006, p 132-134).

Semua subbagian yang teridentifikasi selanjutnya diberi ID. Sama seperti bagian pembukaan, format ID yang diberikan berupa gabungan antara ID dokumen dengan nama subbagian bersangkutan, misalnya subbagian bab dan pasal. Namun, apabila subbagian tersebut berada dalam subbagian lain sebagai induknya, maka format ID nya adalah berupa gabungan ID subbagian induk dengan ID subbagian itu sendiri. Format ID yang diberikan mengikuti urutan pengelompokan materi yang dapat dilihat pada Gambar 2.25.

Pada Gambar 2.25 terlihat bahwa suatu bab dari UU non-perubahan merupakan subbagian induk dari subbagian bagian dan subbagian pasal jika subbagian bagian tidak ada. Sedangkan subbagian bagian adalah subbagian induk dari subbagian paragraf dan subbagian pasal jika subbagian paragraf tidak ada. Demikian juga subbagian paragraf merupakan subbagian induk dari subbagian pasal. Namun, terdapat pengecualian untuk pemberian ID subbagian pasal, yaitu

format ID nya tidak mengikutsertakan ID dari subbagian induknya namun menggunakan ID dari dokumen yang diikuti ID pasal itu sendiri. Hal ini dikarenakan subbagian pasal dalam suatu dokumen UU merupakan subbagian yang unik dengan nomor urut yang berbeda.

Subbagian pasal terdiri dari materi yang dapat dikelompokkan lagi menjadi beberapa subbagian ayat. Selain itu, subbagian pasal juga dapat dikelompokkan menjadi beberapa subbagian huruf atau angka. Dengan demikian, subbagian pasal dapat menjadi subbagian induk dari subbagian ayat, huruf, dan angka. Subbagian ayat juga dapat menjadi subbagian induk dari subbagian huruf dan angka. Sedangkan subbagian huruf dan subbagian angka dapat saling menjadi subbagian induk. Artinya, subbagian huruf dapat dikelompokkan menjadi beberapa subbagian angka. Begitu juga sebaliknya, subbagian angka dapat dikelompokkan menjadi beberapa subbagian huruf.

Jadi, format ID dari subbagian ayat terdiri dari ID subbagian pasal ditambah dengan ID ayat itu sendiri. Sedangkan, ID subbagian huruf dan angka tergantung di subbagian induk mana huruf dan angka tersebut berada seperti yang telah dijelaskan di atas. Format ID nya adalah gabungan antara ID subbagian induk dan ID huruf atau angka itu sendiri. Berikut daftar format ID setiap bagian pada bagian batang tubuh uu non-perubahan:

**Tabel 3.3: Daftar Format ID Subbagian pada Bagian Batang Tubuh UU Non-Perubahan**

SUBBAGIAN	SUBBAGIAN INDUK	FORMAT ID
BAB	-	"[ID_dokumen].bb-[no_bab]"
BAGIAN	BAB	"[ID_bab].bg-[no_bagian]"
PARAGRAF	BAGIAN	"[ID_bagian].pr-[no_paragraf]"
PASAL	BAB, BAGIAN, atau PARAGRAF	"[ID_dokumen].ps-[no_pasal]"
AYAT	PASAL	"[ID_pasal].ay-[no_ayat]"
HURUF	PASAL	"[ID_pasal].hr-[no_huruf]"
	AYAT	"[ID_ayat].hr-[no_huruf]"
	ANGKA	"[ID_angka].hr-[no_huruf]"
ANGKA	PASAL	"[ID_pasal].ag-[no_angka]"
	AYAT	"[ID_ayat].ag-[no_angka]"
	HURUF	"[ID_huruf].ag-[no_angka]"

Pada proses identifikasi ID subbagian juga dilakukan pemberian atribut NO yang merepresentasikan nomor urut subbagian. Seluruh subbagian batang tubuh uu non perubahan diberi atribut tersebut. Isi dari atribut NO diperoleh dari

nomor urut subbagian yang melekat pada *tag* penanda atau dari token yang berada setelah *tag* penanda. Aturan dalam proses identifikasi ID subbagian dalam bagian batang tubuh uu non-perubahan menggunakan aturan yang dikembangkan oleh Mawaddah (2006). Penjelasan selengkapnya dari aturan-aturan tersebut dapat dilihat pada Mawaddah (2006, p 138-139).

Proses selanjutnya adalah pembuatan elemen XML. Pembuatan elemen XML dilakukan dengan cara memberikan *tag* pembuka dan *tag* penutup. *Tag* pembuka subbagian dibuat dengan cara mengubah format *tag* penanda subbagian tersebut menjadi *tag* bernama. Contohnya, *tag* penanda subbagian bab, yaitu <SUB\_BAGIAN TYPE="BAB"> diubah menjadi *tag* bernama <BAB>.

Isi elemen dibuat dengan memperhatikan karakteristik setiap subbagian yang berada dalam bagian. Jika suatu subbagian tidak memiliki subbagian anak di dalamnya maka isi elemen XML nya adalah berupa gabungan token-token yang berada setelah *tag* penandanya sampai token sebelum *tag* penanda berikutnya. Sedangkan untuk subbagian yang memiliki subbagian anak proses pembuatan isi elemennya lebih rumit. Seperti subbagian bab dapat memiliki subbagian anak berupa subbagian bagian atau subbagian pasal. Selain itu subbagian bab juga memiliki judul yang juga dibuat elemen XML nya. Dengan demikian isi dari elemen XML subbagian bab terdiri dari elemen JUDUL\_BAB, elemen BAGIAN jika terdapat subbagian bagian, atau elemen PASAL.

Pada bagian batang tubuh uu non-perubahan dapat ditemukan frase-frase pengacuan, yaitu pada subbagian pasal, ayat, huruf, dan angka. Oleh karena itu, isi elemen subbagian tersebut harus diproses kembali melalui proses pembuatan referensi. Proses pembuatan referensi ini menerima masukan berupa isi elemen, ID dokumen, ID subbagian angka, huruf, ayat, pasal, dan bab. Satu atau beberapa ID yang dikirim tersebut dapat saja bernilai *null* atau kosong sesuai dengan posisi subbagian yang diproses dalam batang tubuh. Jika pada SEI yang dikembangkan Mawaddah (2006) ID subbagian bab tidak dikirimkan, maka pada SEI yang dikembangkan pada penelitian ini ID subbagian bab dikirimkan ke proses pembuatan referensi. ID bab ini diperlukan untuk pembuatan referensi yang mengacu pada subbagian bab dan bagian baik dalam dokumen yang sama atau



dokumen lain yang belum bisa ditangani oleh SEI yang dikembangkan Mawaddah (2006).

Isi elemen XML yang telah selesai diproses oleh proses pembuatan referensi dikirimkan kembali ke proses pembuatan elemen XML. Selanjutnya elemen-elemen subbagian tersebut diberi *tag* penutup yang bersesuaian dengan *tag* pembukanya. Contohnya, elemen XML subbagian bab dengan *tag* pembuka <BAB> diberi *tag* penutup </BAB>. Kemudian, seluruh elemen yang terletak dalam bagian batang tubuh UU non-perubahan diberi elemen induk dengan cara meletakkan seluruh elemen tersebut di antara *tag* <BATANG\_TUBUH> dan </BATANG\_TUBUH>. Aturan pembuatan elemen XML bagian batang tubuh UU non-perubahan pada penelitian ini menggunakan aturan yang dikembangkan Mawaddah (2006). Penjelasan selengkapnya mengenai aturan-aturan tersebut dapat dilihat pada Mawaddah (2006, p 146-149).

### 3.11 Proses Bagian Batang Tubuh UU Non-Perubahan

Batang tubuh UU perubahan terdiri dari beberapa subbagian seperti yang terlihat pada tabel di bawah ini:

Tabel 3.4: Daftar Subbagian pada Bagian Batang Tubuh UU Perubahan

Subbagian	Keterangan	Karakteristik
PASAL_INTI	Rumusan materi dari UU yang mengubah	<ul style="list-style-type: none"> <li>• Diawali kata 'Pasal' yang diikuti nomor urut berupa angka Romawi (I, II, III, dan seterusnya)</li> <li>• Sebelum kata 'Pasal' tidak boleh berupa kata hubung atau kata yang diakhiri tanda baca koma</li> <li>• Setelah nomor urut tidak boleh berupa kata yang diakhiri tanda baca titik dua</li> </ul>
ANGKA_INTI	Rincian materi PASAL_INTI berupa materi-materi yang diubah	<ul style="list-style-type: none"> <li>• Ditandai dengan angka Arab yang diikuti tanda baca titik (1., 2., 3., dan seterusnya)</li> <li>• Sebelum angka Arab tidak boleh berupa kata 'angka' atau 'pasal'</li> <li>• Ditandai kata atau frase awal berupa:               <ol style="list-style-type: none"> <li>1. 'Ketentuan Pasal ...'</li> <li>2. 'Dengan berlakunya Undang-Undang tentang Perubahan ...'</li> <li>3. 'Pasal/Bab/ayat ...'</li> <li>4. 'Di antara Pasal/ayat/Bab/Bagian ... dan Pasal/Bagian/Paragraf/Bab/ayat/Bagian ...' (di beberapa dokumen, penulisan kata 'di' dan 'antara'</li> </ol> </li> </ul>

		<p>digabung)</p> <ol style="list-style-type: none"> <li>5. 'Dalam Bab ...'</li> <li>6. 'Penjelasan Umum/Pasal/Bab ...'</li> <li>7. 'Judul/Sebelum Bagian ...'</li> <li>8. 'Menambah ketentuan/Bab baru ...'</li> <li>9. 'Di antara Pasal ... Bab ... Bagian ... Paragraf ...'</li> <li>10. 'Menambah dua ketentuan baru ...'</li> <li>11. 'Judul dan isi Bab ...'</li> <li>12. 'Ketentuan dalam Pasal ...'</li> </ol>
BAB	Bab UU yang diubah	<ul style="list-style-type: none"> <li>• Diawali kata 'BAB' atau "'BAB' yang diikuti nomor urut berupa angka romawi atau angka romawi yang diikuti huruf kapital, baik terpisah atau tersambung (I, IA, I A, dan seterusnya)</li> <li>• Sebelum kata 'BAB' atau "'BAB' tidak boleh berupa kata hubung atau kata yang diakhiri tanda baca koma</li> <li>• Sebelum kata 'BAB' atau "'BAB' harus berupa kata yang diakhiri tanda baca titik dua</li> <li>• Setelah nomor urut tidak boleh berupa kata yang diakhiri tanda baca titik dua</li> </ul>
PASAL_UBAH	Rumusan materi UU yang diubah	<ul style="list-style-type: none"> <li>• Diawali kata 'Pasal' atau "'Pasal' yang diikuti nomor urut berupa angka Arab atau angka Arab yang diikuti huruf kapital, baik terpisah atau tersambung (1, 1A, 1 A, dan seterusnya)</li> <li>• Sebelum kata 'Pasal' atau "'Pasal' tidak boleh berupa kata hubung atau kata yang diakhiri tanda baca koma</li> <li>• Sebelum kata 'Pasal' atau "'Pasal' harus berupa kata yang diakhiri tanda baca titik dua, tanda baca petik, atau tanda baca titik, serta dapat juga berupa kata 'berikut' atau kata yang ditulis dengan huruf kapital</li> <li>• Setelah nomor urut tidak boleh berupa kata yang diakhiri tanda baca titik dua</li> </ul>
AYAT	Rincian materi PASAL_INTI dan PASAL_UBAH	<ul style="list-style-type: none"> <li>• Ditandai dengan adanya angka Arab atau angka Arab yang diikuti huruf kecil yang diapit tanda baca kurung ((1), (1a), dan seterusnya)</li> <li>• Sebelum angka Arab tidak boleh berupa kata 'ayat' atau kata hubung</li> </ul>
ANGKA_PASAL	Pokok pikiran dari subbagian PASAL_INTI, PASAL_UBAH, AYAT, atau HURUF UU yang diubah	<ul style="list-style-type: none"> <li>• Ditandai dengan adanya angka Arab yang diikuti tanda baca titik (1., 2., 3., dan seterusnya)</li> <li>• Sebelum angka Arab tidak boleh berupa kata 'angka' atau 'pasal'</li> </ul>
HURUF	Pokok	<ul style="list-style-type: none"> <li>• Ditandai dengan adanya huruf abjad</li> </ul>

	pikiran dari PASAL_INTI, PASAL_UBAH, AYAT, atau ANGKA_PASAL	yang diikuti tanda baca titik (a., b., c., dan seterusnya) <ul style="list-style-type: none"> <li>• Sebelum huruf abjad tidak boleh berupa kata 'huruf'</li> </ul>
HURUF_RINCI	Pokok pikiran dari subbagian ANGKA_PASAL yang berada dalam subbagian HURUF_UU yang diubah	<ul style="list-style-type: none"> <li>• Ditandai dengan adanya huruf abjad yang diikuti tanda baca kurung tutup (a), b), c), dan seterusnya)</li> <li>• Sebelum huruf abjad tidak boleh berupa kata 'huruf'</li> </ul>

Berdasarkan karakteristik yang dimiliki masing-masing subbagian maka dapat dibuat aturan-aturan untuk mengidentifikasi setiap subbagian tersebut. Aturan-aturan tersebut dapat dilihat pada Lampiran 9.

Seluruh subbagian yang teridentifikasi selanjutnya diberi ID. ID yang diberikan pada suatu subbagian memiliki format berupa gabungan antara ID dokumen dengan nama subbagian bersangkutan. Apabila subbagian tersebut berada dalam subbagian lain sebagai induknya, maka format ID nya adalah berupa gabungan ID subbagian induk dengan ID subbagian itu sendiri. Format ID mengikuti urutan pengelompokan materi yang dapat dilihat pada Gambar 2.38.

Pada Gambar 2.38 terlihat bahwa subbagian pasal inti merupakan subbagian induk dari subbagian angka inti dan subbagian pasal yang diubah jika subbagian angka inti tidak ada. Sedangkan subbagian angka inti menjadi subbagian induk dari bab atau pasal yang diubah. Bab yang diubah menjadi subbagian induk dari subbagian lain sama seperti halnya bab pada UU non-perubahan.

ID dari subbagian angka inti menggunakan ID subbagian induknya, yakni subbagian pasal inti, diikuti ID angka inti itu sendiri. Namun demikian, pemberian ID untuk materi yang diubah, seperti bab dan pasal, tidak tergantung pada ID subbagian induknya yakni pasal inti atau angka inti. Format ID nya menggunakan ID dokumen UU yang diubah diikuti ID subbagian itu sendiri. Dan sama halnya dengan pasal dari UU non-perubahan, pasal yang diubah pada UU perubahan juga tidak menggunakan ID subbagian induknya, tapi menggunakan ID dokumen UU yang diubah.

Subbagian pasal yang diubah pada UU perubahan terdiri dari materi yang dapat dikelompokkan lagi menjadi beberapa subbagian ayat. Selain itu, subbagian pasal juga dapat dikelompokkan menjadi beberapa subbagian huruf atau angka. Dengan demikian, subbagian pasal dapat menjadi subbagian induk dari subbagian ayat, huruf, dan angka.

Subbagian ayat juga dapat menjadi subbagian induk dari subbagian huruf dan angka. Sedangkan subbagian huruf dan subbagian angka dapat saling menjadi subbagian induk. Artinya, subbagian huruf dapat dikelompokkan menjadi beberapa subbagian angka. Begitu juga sebaliknya, subbagian angka dapat dikelompokkan menjadi beberapa subbagian huruf.

Jadi, format ID dari subbagian ayat terdiri dari ID subbagian pasal ditambah dengan ID ayat itu sendiri. Sedangkan, ID subbagian huruf dan angka tergantung di subbagian induk mana huruf dan angka tersebut berada seperti yang telah dijelaskan di atas. Format ID nya adalah gabungan antara ID subbagian induk dan ID huruf atau angka itu sendiri. Berikut daftar format ID setiap bagian pada bagian batang tubuh UU perubahan:

**Tabel 3.5: Daftar Format ID Subbagian pada Bagian Batang Tubuh UU Perubahan**

SUBBAGIAN	SUBBAGIAN INDUK	FORMAT ID
PASAL_INTI	-	"[ID_dokumen].ps-[no_pasal_inti]"
ANGKA_INTI	PASAL_INTI	"[ID_pasal_inti].ag-[no_angka_inti]"
BAB	ANGKA_INTI	"[ID_dokumen_diubah].bb-[no_bab]"
PASAL_UBAH	PASAL_INTI, ANGKA_INTI, atau BAB	"[ID_dokumen_diubah].ps-[no_pasal_ubah]"
AYAT	PASAL_INTI	"[ID_pasal_inti].ay-[no_ayat]"
	PASAL_UBAH	"[ID_pasal_ubah].ay-[no_ayat]"
ANGKA_PASAL	PASAL_INTI	"[ID_pasal_inti].ag-[no_angka]"
	PASAL_UBAH	"[ID_pasal_ubah].ag-[no_angka]"
	AYAT	"[ID_ayat].ag-[no_angka]"
	HURUF	"[ID_huruf].ag-[no_angka]"
HURUF	PASAL_INTI	"[ID_pasal_inti].hr-[no_huruf]"
	PASAL_UBAH	"[ID_pasal_ubah].hr-[no_huruf]"
	AYAT	"[ID_ayat].hr-[no_huruf]"
	ANGKA_PASAL	"[ID_angka_pasal].hr-[no_huruf]"
HURUF_RINCI	ANGKA_PASAL	"[ID_angka_pasal].hr-[no_huruf]"

Pada proses identifikasi ID subbagian juga dilakukan pemberian atribut NO yang merepresentasikan nomor urut subbagian. Seluruh subbagian batang tubuh uu perubahan diberi atribut tersebut. Isi dari atribut NO diperoleh dari nomor urut subbagian yang melekat pada *tag* penanda atau dari token yang berada setelah *tag* penanda. Aturan dalam proses identifikasi ID dan pemberian atribut NO pada subbagian dalam bagian batang tubuh uu perubahan dapat dilihat pada Lampiran 10.

Proses selanjutnya adalah pembuatan elemen XML. Elemen XML terdiri dari *tag* pembuka, isi elemen, dan *tag* penutup. *Tag* pembuka subbagian dibuat dengan cara mengubah format *tag* penanda subbagian tersebut menjadi *tag* bernama. Sedangkan, isi elemen dibuat dengan memperhatikan karakteristik setiap subbagian yang berada dalam bagian. Jika suatu subbagian tidak memiliki subbagian anak di dalamnya maka isi elemen XML nya adalah berupa gabungan token-token yang berada setelah *tag* penandanya sampai token sebelum *tag* penanda berikutnya. Sedangkan, untuk subbagian yang memiliki subbagian anak proses pembuatan isi elemennya lebih rumit. Seperti subbagian pasal inti dapat memiliki subbagian anak berupa subbagian angka inti atau subbagian pasal yang diubah. Dengan demikian isi dari elemen XML subbagian pasal inti terdiri dari elemen ANGKA jika terdapat subbagian angka inti, atau elemen PASAL.

Frase pengacuan dapat ditemukan pada subbagian pasal inti, angka inti, pasal ubah, ayat, huruf, angka pasal, dan huruf rinci. Oleh karena itu, isi elemen subbagian tersebut harus diproses kembali melalui proses pembuatan referensi. Proses tersebut akan menerima masukan berupa isi elemen beserta ID dokumen, ID dokumen yang diubah, ID subbagian angka, huruf, ayat, pasal, dan bab. ID dokumen yang diubah ini diperlukan untuk pembuatan referensi pada subbagian pasal ubah, ayat, angka pasal, huruf, dan huruf rinci.

Isi elemen XML yang telah selesai diproses oleh proses pembuatan referensi dikirimkan kembali ke proses pembuatan elemen XML. Selanjutnya elemen-elemen subbagian tersebut diberi *tag* penutup yang bersesuaian dengan *tag* pembukanya. Kemudian, seluruh elemen pada bagian batang tubuh UU perubahan diberi elemen induk dengan cara meletakkannya di antara *tag* pembuka <BATANG\_TUBUH> dan *tag* penutup </BATANG\_TUBUH>.

Aturan pembuatan elemen XML setiap subbagian pada bagian batang tubuh UU perubahan dapat dilihat pada Lampiran 11.

### 3.12 Proses Bagian Penutup

Bagian penutup terdiri dari sebelas subbagian. Pada penelitian ini ditemukan satu subbagian lagi yaitu subbagian ditetapkan yang tidak ditemukan pada penelitian Mawaddah (2006). Subbagian ditetapkan hadir bergantian dengan subbagian disahkan, artinya suatu dokumen dapat disahkan atau ditetapkan. Penjelasan dari masing-masing subbagian tersebut dapat dilihat pada tabel di bawah ini:

**Tabel 3.6: Daftar Subbagian pada Bagian Penutup**

Subbagian	Keterangan	Karakteristik
RUMUSAN_PERINTAH	Rumusan perintah pengundangan dan penempatan UU dalam Lembaran Negara Republik Indonesia	Ditandai dengan frase 'Agar setiap/semua orang mengetahuinya, memerintahkan pengundangan ...' atau 'Agar setiap/semua orang dapat mengetahuinya, memerintahkan pengundangan ...'
DISAHKAN	Penandatanganan pengesahan UU yang terdiri dari tempat dan tanggal pengesahan, nama jabatan, dan nama lengkap pejabat yang mengesahkan	Ditandai dengan kata 'Disahkan'
DITETAPKAN	<b>Penandatanganan penetapan UU yang terdiri dari tempat dan tanggal penetapan, nama jabatan, dan nama lengkap pejabat yang menetapkan.</b>	Ditandai dengan kata 'Ditetapkan'
DIUNDANGKAN	Pengundangan UU yang terdiri dari tempat dan tanggal pengundangan, nama jabatan, dan nama lengkap pejabat yang mengundangan	Ditandai dengan kata 'Diundangkan'
LOKASI	Tempat pengesahan, penetapan, dan pengundangan	Ditandai dengan kata hubung 'di'
WAKTU	Waktu pengesahan, penetapan, dan pengundangan yang	Ditandai dengan frase 'pada tanggal'

	terdiri dari tanggal, bulan, dan tahun	
POSISI	Posisi atau jabatan pelaku pengesahan, penetapan, dan pengundangan	Ditandai dengan gabungan kata-kata yang terdapat dalam kamus nama jabatan yang dapat membentuk frase nama jabatan
NAMA	Nama lengkap pejabat yang mengesahkan, menetapkan, dan mengundang	Ditandai dengan gabungan kata-kata yang terdapat dalam kamus nama pejabat yang dapat membentuk nama lengkap pejabat
LEMBARAN	Petunjuk pencatuman UU dalam Lembaran Negara yang terdiri dari nomor dan tahun lembaran	Ditandai dengan kata 'Lembaran'
TAHUN	Tahun dari Lembaran Negara	Ditandai dengan kata 'tahun' yang diikuti empat digit angka Arab
NOMOR	Nomor dari Lembaran Negara	Ditandai dengan kata 'nomor' yang diikuti angka Arab (1, 2, 3, dan seterusnya)

Keterangan: Bagian yang dicetak hitam tebal adalah bagian yang berbeda dengan penelitian yang dilakukan Mawaddah (2006)

Dari masing-masing karakteristik yang dimiliki setiap subbagian, maka dapat dibuat aturan-aturan untuk mengidentifikasi subbagian tersebut. Aturan-aturan identifikasi subbagian penutup dapat dilihat pada Lampiran 12.

Setelah seluruh subbagian teridentifikasi, proses selanjutnya adalah pembuatan elemen-elemen XML sesuai dengan subbagian yang ditemukan. Elemen XML terdiri dari *tag* pembuka, isi elemen, dan *tag* penutup. *Tag* pembuka subbagian dibuat dengan cara mengubah format *tag* penanda subbagian tersebut menjadi *tag* bernama. Sedangkan, isi elemen dibuat dengan memperhatikan karakteristik setiap subbagian yang berada dalam bagian. Jika suatu subbagian tidak memiliki subbagian anak di dalamnya maka isi elemen XML nya adalah berupa gabungan token-token yang berada setelah *tag* penandanya sampai token sebelum *tag* penanda berikutnya.

Sedangkan untuk subbagian yang memiliki subbagian anak proses pembuatan isi elemennya lebih rumit. Seperti subbagian disahkan yang terdiri dari subbagian lokasi, waktu, dan pelaku dengan subbagian waktu terdiri dari subbagian tanggal, bulan, dan tahun dan subbagian pelaku terdiri dari subbagian

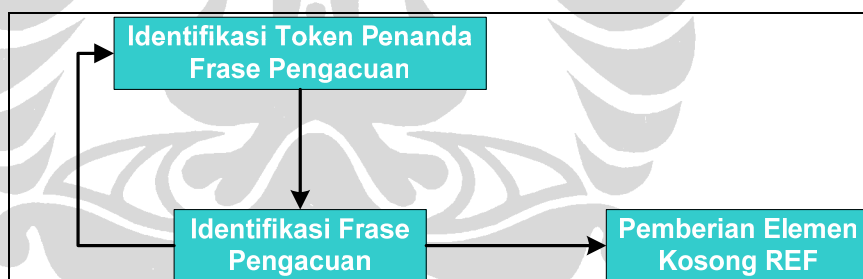
posisi dan nama. Proses terakhir adalah pemberian elemen induk dengan cara meletakkannya di antara *tag* pembuka <PENUTUP> dan *tag* penutup </PENUTUP>.

Aturan pembuatan elemen XML setiap subbagian pada bagian penutup dapat dilihat pada Lampiran 13.

### 3.13 Pembuatan Referensi

Proses pembuatan referensi dipanggil pada saat pemrosesan bagian judul, pembukaan, dan batang tubuh, baik batang tubuh UU non-perubahan maupun batang tubuh UU perubahan. Proses ini menerima masukan berupa isi elemen yang merupakan isi materi suatu subbagian beserta ID subbagian tersebut dan ID subbagian induknya.

Proses pembuatan referensi diawali dengan menguraikan isi elemen menjadi token-token. Token-token tersebut kemudian diproses dengan tiga sub proses, yaitu sub proses identifikasi token penanda frase pengacuan, identifikasi frase pengacuan, dan pemberian elemen kosong REF. Alur dari ketiga sub proses tersebut dapat dilihat pada gambar di bawah ini:



Gambar 3.5: Alur Proses Pembuatan Referensi

Sumber: Mawaddah, S. (2006). Pengembangan standar dokumen legal indonesia berbasis XML menggunakan sistem ekstraksi informasi dengan studi kasus dokumen undang-undang republik Indonesia.

#### 3.13.1 Sub Proses Identifikasi Token Penanda Frase Pengacuan

Proses identifikasi token penanda frase pengacuan berfungsi untuk menentukan awal dari keberadaan frase pengacuan. Identifikasi token dilakukan dengan cara mencocokkan setiap token dari isi elemen dengan token penanda frase pengacuan yang telah didefinisikan. Token yang cocok diproses lebih lanjut, sedangkan token yang tidak cocok disatukan kembali untuk dikirim ke proses yang mengirimnya, yaitu proses bagian judul, pembukaan, atau batang tubuh.



Berdasarkan hasil analisis terhadap bagian judul, pembukaan, dan batang tubuh dokumen observasi, terdapat beberapa jenis pengacuan. Jenis-jenis pengacuan yang ditemukan pada penelitian ini lebih bervariasi dibanding jenis pengacuan yang ditemukan pada penelitian Mawaddah (2006). Berikut daftar pengacuan tersebut:

**Tabel 3.7: Jenis-Jenis Frase Pengacuan Diidentifikasi Sistem**

Jenis Pengacuan	Token Penanda Pengacuan	Contoh
Pengacuan ke UUD 1945	kata 'undang-undang'	berdasarkan Pancasila dan <u>Undang-Undang Dasar Negara Republik Indonesia Tahun 1945</u> , menjamin...
Penegasan nama dokumen yang sedang diproses	kata 'undang-undang'	... sebagaimana dimaksud dalam undang-undang ini.
Pengacuan ke dokumen UU lain	kata 'undang-undang'	Perubahan ketiga atas <u>Undang-Undang Nomor 6 Tahun 1983</u> tentang Ketentuan Umum dan Tata Cara Perpajakan
Pengacuan ke UU Pajak dan Pertambahan Nilai 1984	kata 'undang-undang'	... dengan mengacu pada ketentuan <u>Undang-Undang Pajak Pertambahan Nilai 1984</u> .
Pengacuan ke Tap MPR	kata 'ketetapan'	<u>Ketetapan Majelis Permusyawaratan Rakyat Republik Indonesia Nomor IV/MPR/1999</u> tentang Garis-garis Besar Haluan Negara (GBHN) Tahun 1999 - 2004
Pengacuan ke Kitab Undang-Undang Hukum Pidana (KUHP)	kata 'kitab'	dalam <u>Kitab Undang-undang Hukum Pidana</u> ...
Pengacuan ke subbagian huruf dalam subbagian konsiderans	kata 'huruf' atau 'butir'	<ul style="list-style-type: none"> <li>berdasarkan pertimbangan pada <u>huruf a, huruf b, dan huruf c</u> ...</li> <li>berdasarkan pertimbangan pada <u>butir a, butir b, dan butir c</u> ...</li> </ul>
Pengacuan ke subbagian angka dalam subbagian dasar hukum	kata 'angka'	seperti <u>angka 1</u> .
Pengacuan ke subbagian dalam bagian batang tubuh dokumen UU lain	kata 'pasal'	sebagaimana dimaksud dalam <u>Pasal 74 dan Pasal 123 Undang-Undang Nomor 31 Tahun 1997</u> tentang Peradilan Militer.
Pengacuan ke subbagian dalam bagian batang tubuh dokumen UUD 1945	kata 'pasal'	<u>Pasal 5 ayat (1), Pasal 20, dan Pasal 23A Undang-Undang Dasar Negara Republik Indonesia Tahun</u>

		1945
<b>Pengacuan ke subbagian dalam bagian batang tubuh KUHP</b>	<b>kata 'pasal'</b>	<b><u>Pasal 107 dan Pasal 108 Bab I Buku 11 Kitab Undang-undang Hukum Pidana</u></b>
Pengacuan ke subbagian dalam batang tubuh dokumen yang diproses	kata 'bab', 'bagian', 'pasal', 'ayat', 'huruf', 'butir', atau 'angka'	<ul style="list-style-type: none"> <li>• sesuai dengan <u>angka 7 huruf c</u> ...</li> <li>• sebagaimana dimaksud dalam <u>ayat (1) dan ayat (2)</u> tidak boleh...</li> <li>• seperti dalam <u>Pasal 1, Pasal 3, Pasal 7, Pasal 9, atau Pasal 10</u>.</li> <li>• Di antara <u>Pasal 106 dan BAB VII</u> disisipkan ...</li> <li>• ... dalam Bagian Kedua, Bagian Ketiga, dan Bagian Keempat ...</li> </ul>
Penegasan nama subbagian yang sedang diproses	kata 'bab', 'bagian', 'pasal', 'ayat', 'huruf', 'butir', atau 'angka'	<ul style="list-style-type: none"> <li>• sebagaimana dimaksud dalam <u>pasal ini</u>.</li> <li>• ... dalam <u>ayat ini</u> ...</li> <li>• ... dalam <u>bab ini</u> ...</li> </ul>
<b>Pengacuan ke subbagian dalam UU Perbendaharaan Indonesia (Indische Comptabiliteitswet)</b>	<b>kata 'pasal'</b>	<b><u>... Perubahan Pasal 7 Indische Comptabiliteitswet ...</u></b>

Keterangan: Bagian yang dicetak hitam tebal adalah bagian yang berbeda dengan penelitian yang dilakukan Mawaddah (2006)

Berdasarkan Tabel 3.7 di atas, maka token penanda adanya frase pengacuan adalah berupa token 'undang-undang', 'ketetapan', 'kitab', 'bab', 'bagian', 'pasal', 'ayat', 'angka', 'huruf', dan 'butir'. Setelah token penanda pengacuan ditemukan, token tersebut beserta indeksnya dikirim ke sub proses selanjutnya, yaitu sub proses identifikasi frase pengacuan untuk mendapatkan frase-frase pengacuan yang mungkin ada.

### 3.13.2 Sub Proses Identifikasi Frase Pengacuan

Token penanda yang telah ditemukan pada sub proses sebelumnya menjadi kata awal dari frase pengacuan. Untuk menentukan batas akhirnya maka dilakukan identifikasi frase pengacuan yang cocok dengan pola pengacuan yang telah didefinisikan. Daftar pola pengacuan dari setiap token penanda pengacuan dapat dilihat pada Tabel 3.8.

Tabel 3.8: Pola Pengacuan dari Setiap Token Penanda Pengacuan

Token Penanda Pengacuan	Contoh Pola Pengacuan
kata 'undang-undang'	<ul style="list-style-type: none"> <li>• Undang-Undang ini</li> <li>• Undang-Undang Dasar 1945</li> <li>• Undang-Undang Dasar Negara Republik Indonesia Tahun 1945</li> <li>• Undang-Undang Nomor 3 Tahun 1998</li> <li>• Undang-Undang No.16 Tahun 2001</li> <li>• Undang-undang No. 1 Tahun 1979</li> <li>• Undang-Undang Republik Indonesia Nomor 4 Tahun 1970</li> <li>• Undang-Undang Pajak Pertambahan Nilai 1984</li> </ul>
kata 'ketetapan'	<ul style="list-style-type: none"> <li>• Ketetapan Majelis Permusyawaratan Rakyat Republik Indonesia Nomor X/MPR/1998</li> <li>• Ketetapan Majelis Permusyawaratan Rakyat Sementara Republik Indonesia No.XXV/MPRS/1966</li> </ul>
kata 'kitab'	Kitab Undang-Undang Hukum Pidana
kata 'bab'	<ul style="list-style-type: none"> <li>• Bab ini</li> <li>• Bab IX</li> <li>• Bab IX dan Bab X</li> </ul>
kata 'bagian'	<ul style="list-style-type: none"> <li>• Bagian Pertama Bab IX</li> <li>• Bagian Pertama, ..., dan Bagian Ketiga Bab X</li> </ul>
kata 'pasal'	<ul style="list-style-type: none"> <li>• Pasal ini</li> <li>• Pasal 10</li> <li>• Pasal 10 ayat (1)</li> <li>• Pasal 1 angka 3</li> <li>• Pasal 7 huruf a</li> <li>• Pasal 10 [kata hubung] Pasal 15</li> <li>• Pasal 1 ayat (1) huruf c</li> <li>• Pasal 1 ayat (1) angka 2</li> <li>• Pasal 1 huruf a angka 1</li> <li>• Pasal 1, Pasal 2, ... [kata hubung] Pasal 5</li> <li>• Pasal 8 huruf a, b, ..., [kata hubung] g</li> <li>• Pasal 1 ayat (1) huruf a angka 1</li> <li>• Pasal 2 ayat (2) angka 1 huruf a</li> <li>• Pasal 21 ayat (1) dan (3)</li> <li>• Pasal 21 ayat (4), ..., [kata hubung] Pasal 23 ayat (3)</li> <li>• Pasal 28 ayat (1) [kata hubung] ayat (3), ... [kata hubung] Pasal 30</li> <li>• Pasal 1 ayat (1) huruf a, Pasal 2 ayat (2) angka 3, Pasal 4 ayat (1) [kata hubung] ayat (3), ..., [kata hubung] Pasal 2 huruf a angka 3</li> <li>• Pasal 1, ..., [kata hubung] Pasal 3 Undang-Undang ini</li> <li>• Pasal 37 Undang-Undang Dasar 1945</li> <li>• Pasal 5 ayat (1), Pasal 11, ..., [kata hubung] Pasal 37 Undang-Undang Dasar Negara Republik Indonesia Tahun 1945</li> <li>• Pasal 104 ayat (1) Undang-undang Nomor 39 Tahun 1999</li> </ul>

	<ul style="list-style-type: none"> <li>• Pasal 74, ..., [kata hubung] Pasal 123 Undang-undang Nomor 31 Tahun 1997</li> <li>• Pasal 107 Kitab Undang-Undang Hukum Pidana</li> <li>• Pasal 107 A, ..., dan Pasal 107 F Kitab Undang-Undang Hukum Pidana</li> <li>• Pasal 7 Indische Comptabiliteitswet</li> </ul>
kata 'ayat'	<ul style="list-style-type: none"> <li>• ayat ini</li> <li>• ayat (1)</li> <li>• ayat (1) [kata hubung] ayat (2)</li> <li>• ayat (1), ayat (2), ..., [kata hubung] ayat (4)</li> <li>• ayat (1) huruf a</li> <li>• ayat (1) angka 1</li> <li>• ayat (1) huruf b angka 3</li> <li>• ayat (1) huruf a, huruf b, ..., dan huruf [no huruf]</li> </ul>
kata 'angka'	<ul style="list-style-type: none"> <li>• angka ini</li> <li>• angka 1</li> <li>• angka 1 [kata hubung] angka 3</li> <li>• angka 5 huruf a</li> <li>• angka 1, ..., [kata hubung] angka 4</li> </ul>
kata 'huruf'	<ul style="list-style-type: none"> <li>• huruf ini</li> <li>• huruf a</li> <li>• huruf a [kata hubung] huruf b</li> <li>• huruf a, b, ..., dan f</li> <li>• huruf a, huruf b, ..., [kata hubung] huruf f</li> <li>• huruf b angka 4</li> </ul>
kata 'butir'	<ul style="list-style-type: none"> <li>• butir ini</li> <li>• butir a</li> <li>• butir a [kata hubung] butir b</li> <li>• butir a, b, ..., dan f</li> <li>• butir a, butir b, ..., [kata hubung] butir f</li> <li>• butir b angka 4</li> </ul>

Pada tabel di atas terdapat tulisan 'kata hubung' yang diletakkan di dalam tanda kurung siku ([kata hubung]). Tulisan tersebut menunjukkan adanya kata hubung yang terdapat di antara frase-frase pengacuan. Kata hubung tersebut dapat berupa kata 'dan', 'atau', 'dan/atau', 'kecuali', 'selain', dan 'serta'. Selain itu dapat juga berupa frase 'sampai dengan'.

Berikut contoh proses identifikasi token penanda frase pengacuan dan identifikasi frase pengacuan:

<p><u>Sub Proses Identifikasi Token Penanda Frase Pengacuan:</u></p> <p>Masukan:  Pemberhentian anggota DPRA/DPRK sebagaimana dimaksud pada ayat (2) huruf b, huruf c, huruf d, huruf e, dan huruf f dilaksanakan setelah ada keputusan DPRA/DPRK berdasarkan rekomendasi dari Badan Kehormatan DPRA/DPRK.</p>
--

Isi Elemen Awal:  
 Pemberhentian anggota DPRA/DPRK sebagaimana dimaksud pada  
 Token Penanda:  
 ayat

Sub Proses Identifikasi Frase Pengacuan:  
 Token Penanda:  
 ayat  
 Frase Pengacuan:  
 ayat (2) huruf b, huruf c, huruf d, huruf e, dan huruf f  
 Indeks Token yang Dikembalikan ke Sub Proses Identifikasi Token  
 Penanda Frase Pengacuan:  
 Indeks dari token "dilaksanakan"

Sub Proses Identifikasi Token Penanda Frase Pengacuan:  
 Masukan:  
 dilaksanakan setelah ada keputusan DPRA/DPRK berdasarkan  
 rekomendasi dari Badan Kehormatan DPRA/DPRK.  
 Isi Elemen Awal:  
 dilaksanakan setelah ada keputusan DPRA/DPRK berdasarkan  
 rekomendasi dari Badan Kehormatan DPRA/DPRK.  
 Token Penanda: -

**Gambar 3.6: Contoh Identifikasi Token Penanda Frase Pengacuan dan Identifikasi Frase Pengacuan**

Aturan identifikasi frase pengacuan dapat dilihat pada Lampiran 14.

### 3.13.3 Sub Proses Pemberian Elemen Kosong REF

Frase pengacuan yang diperoleh pada sub proses sebelumnya diproses lebih lanjut menjadi sejumlah elemen kosong REF. Elemen tersebut memiliki atribut ID dari subbagian atau dokumen yang menjadi target pengacuan. Langkah awal dari proses pemberian elemen kosong REF adalah pengecekan terhadap penulisan frase pengacuan. Hal ini diperlukan karena terdapat penulisan frase pengacuan yang tidak sesuai dengan aturan yang telah ditetapkan oleh Undang-Undang Nomor 10 Tahun 2004 sehingga diperlukan pembenaran penulisan. Frase pengacuan yang telah dibenarkan penulisannya akan memudahkan proses pemberian elemen kosong REF nya.

Berdasarkan hasil analisis terhadap dokumen observasi, kesalahan penulisan frase pengacuan terdapat pada:

#### 1. Pengacuan terhadap beberapa subbagian huruf/butir

Penulisan pengacuan yang benar adalah penulisan kata 'huruf' di depan setiap nomor subbagian huruf yang menjadi target pengacuan. Contohnya 'huruf a, huruf b, ..., [kata hubung] huruf g'. Namun di beberapa dokumen, penulisan

kata ‘huruf’ hanya pada kata pertama frase atau di depan nomor huruf yang pertama, seperti ‘huruf a, b, [kata hubung], g’. Untuk itu perlu dilakukan membenaran dengan cara menambahkan kata ‘huruf’ di depan setiap nomor huruf.

2. Pengacuan terhadap beberapa subbagian ayat

Penulisan pengacuan yang benar adalah penulisan kata ‘ayat’ di depan setiap nomor subbagian ayat yang menjadi target pengacuan. Contohnya, ‘ayat (1), ayat (2), ..., [kata hubung] ayat (5)’. Namun kesalahan penulisan terjadi di beberapa dokumen di mana kata ‘ayat’ hanya terdapat pada kata pertama frase atau di depan nomor ayat pertama, misalnya ayat (1), (2), ..., [kata hubung] (4). membenaran penulisan dapat dilakukan dengan cara menambahkan kata ‘ayat’ di depan setiap nomor ayat.

3. Pengacuan terhadap beberapa subbagian angka

Penulisan pengacuan yang benar adalah penulisan kata ‘angka’ di depan setiap nomor subbagian angka yang menjadi target pengacuan. Contohnya, ‘angka 1, angka 2, ..., [kata hubung] angka 6’. Namun di beberapa dokumen, penulisan kata ‘angka’ hanya pada kata pertama frase atau di depan nomor angka yang pertama, seperti ‘angka 1, 2, ..., [kata hubung] 5’. Untuk itu perlu dilakukan membenaran dengan cara menambahkan kata ‘angka’ di depan setiap nomor angka.

Setelah penulisan frase pengacuan dinyatakan benar, maka proses pemberian elemen kosong REF dapat dilakukan. Frase pengacuan yang ditangani terlebih dahulu adalah frase-frase dengan target pengacuan dokumen legal lain dan dokumen yang sedang diproses, yaitu frase pengacuan terhadap Tap MPR, UUD 1945, dokumen yang diproses, dokumen UU lain, dan dokumen KUHP. Format elemen REF dari kelima jenis frase tersebut adalah sebagai berikut:

**Tabel 3.9: Format Elemen REF untuk Frase Pengacuan dengan Target Pengacuan Dokumen Lain dan Dokumen yang Diproses**

Frases Pengacuan	Format Elemen REF	Contoh
Tap MPR	<REF ID="tap.[nomor_tap]"/>	Masukan: Ketetapan Majelis Permusyawaratan Rakyat Republik Indonesia Nomor X/MPR/1998 Keluaran:

		<REF ID="tap.X/MPR/1998" />
UUD 1945	<REF ID="uud45" />	Masukan: Undang-Undang Dasar Negara Republik Indonesia Tahun 1945 Keluaran: <REF ID="uud45" />
Dokumen yang diproses	<REF ID="[ ID_dokumen]" />	Masukan: Undang-Undang ini Keluaran: <REF ID="uu-12-2008" />
Dokumen UU lain	<REF ID="uu-[nomor_uu]-[tahun_uu]" />	Masukan: Undang-Undang Nomor 3 Tahun 1998 Keluaran: <REF ID="uu-3-1998" />
Dokumen UU Pajak Pertambahan Nilai 1984	<REF ID=" uuPPN84" />	Masukan: Undang-Undang Pajak Pertambahan Nilai 1984 Keluaran: <REF ID="uuPPN84" />
Dokumen KUHP	<REF ID="kuhp" />	Masukan: Kitab Undang-Undang Hukum Pidana Keluaran: <REF ID="kuhp" />

Frase-frase pengacuan lain yang tidak termasuk pada kelima jenis frase pengacuan pada tabel di atas diproses sesuai dengan jumlah kata atau token yang terdapat dalam frase tersebut. Aturan pemberian elemen REF untuk frase-frase tersebut dapat dilihat pada Lampiran 15.

### 3.14 Pembuatan Dokumen dalam Format XML

Proses terakhir dari fitur standardisasi dokumen adalah proses pembuatan dokumen dalam format XML. Dalam proses ini elemen-elemen XML yang dihasilkan pada proses setiap bagian disatukan sehingga membentuk dokumen UU dalam format XML. Seluruh elemen tersebut disatukan dalam satu elemen akar yaitu elemen akar DOKUMEN.

Elemen akar DOKUMEN dibuka dengan *tag* pembuka <DOKUMEN> dan ditutup dengan *tag* penutup </DOKUMEN>. *Tag* pembuka tersebut memiliki atribut TIPE dan ID. Atribut TIPE berisi tipe dari dokumen yang diproses, dalam hal ini adalah "UU". Sedangkan isi dari atribut ID adalah ID dari dokumen yang diproses. Contohnya, untuk dokumen Undang-Undang Nomor 12 Tahun 2008, *tag* pembukanya adalah <DOKUMEN TIPE="UU" ID="uu-12-2008">.

Untuk menentukan kevalidan dokumen UU dalam format XML yang dihasilkan diperlukan DTD (*Document Type Declaration*). DTD dibuat berdasarkan karakteristik dan penempatan subbagian dari setiap bagian UU. Informasi keberadaan DTD dicantumkan di setiap bagian awal dokumen dengan cara menambahkan *tag* `<!DOCTYPE DOKUMEN SYSTEM "DTDUURI.dtd">`. DTD yang digunakan pada dokumen XML yang dihasilkan dapat dilihat pada Lampiran 16.

### 3.15 Pencarian Referensi

Dokumen XML yang dihasilkan oleh fitur standardisasi dokumen akan menjadi input dari fitur pembuatan graf referensi dokumen. Fitur tersebut terdiri atas dua proses, yaitu proses pencarian referensi dan proses pembuatan dokumen DOT. Pada subbab ini akan dibahas proses yang pertama, yakni proses pencarian referensi.

Proses pencarian referensi diawali dengan penguraian elemen XML menjadi token-token. Token yang dihasilkan berupa elemen-elemen XML tanpa tanda baca '`<`' dan '`>`' pada *tag* pembuka dan *tag* penutupnya. Setelah itu dilakukan pengambilan ID dari dokumen yang sedang diproses. ID diambil dari isi atribut ID yang terdapat pada elemen akar DOKUMEN. Sebagai *node* pertama dari graf dibuat *node* dengan nama sesuai ID yang diperoleh tersebut. *Node* ini akan memiliki hubungan dengan *node* lain yang menggambarkan keterkaitan dokumen yang diproses dengan dokumen lain.

*Node* dari graf dibuat berdasarkan elemen kosong REF yang ditemukan dalam dokumen yang sedang diproses. Pencarian elemen REF dilakukan dengan cara mencari token yang diawali dengan kata 'REF ID='. Selanjutnya, proses substring dilakukan pada setiap token yang cocok sehingga didapatkan isi atribut ID. Setiap ID yang ditemukan dibuatkan *node* nya dan *node* tersebut memiliki hubungan dengan *node* yang bernama sesuai ID dokumen yang diproses.

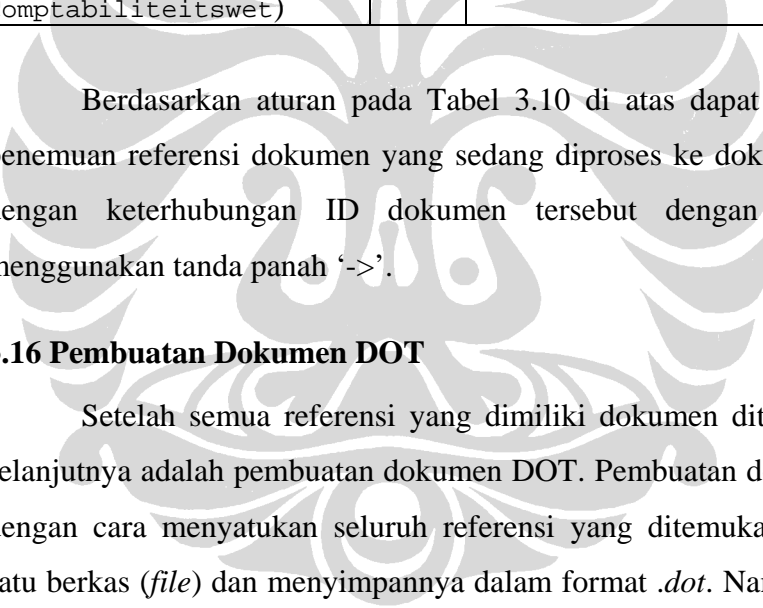
Target referensi suatu dokumen dapat berupa referensi terhadap dokumen UUD 1945, Tap MPR, UU lain, UU yang sedang diproses itu sendiri, KUHP, UU Pajak Pertambahan Nilai 1984, dan UU Perbendaharaan Indonesia (*Indische Comptabiliteitswet*). Selain itu referensi juga dapat terjadi terhadap subbagian-subbagian dari masing-masing dokumen tersebut. Namun graf referensi yang akan



dibuat pada penelitian ini terbatas pada referensi terhadap dokumen saja dan tidak termasuk referensi terhadap subbagian–subbagian di dalamnya. Berikut aturan pencarian ID dari setiap dokumen yang menjadi target referensi tersebut:

**Tabel 3.10: Aturan Pencarian Referensi**

Target Referensi	Aturan	
	Token[i]=Token[i].SubstringSetelah("REF ID=")	
UUD 1945	<b>JIKA</b>	Token[i].Start="uud45"
	<b>MAKA</b>	referensi="[ID_dokumen]->uud45"
Tap MPR	<b>JIKA</b>	Token[i].Start="tap"
	<b>MAKA</b>	referensi="[ID_dokumen]->[ID_tap]"
UU lain dan UU diproses	<b>JIKA</b>	Token[i].Start="uu-"
	<b>MAKA</b>	referensi="[ID_dokumen]->[ID_uu]"
KUHP	<b>JIKA</b>	Token[i].Start="kuhp"
	<b>MAKA</b>	referensi="[ID_dokumen]->kuhp"
UU Pajak Pertambahan Nilai 1984	<b>JIKA</b>	Token[i]="uuPPN84"
	<b>MAKA</b>	referensi="[ID_dokumen]->uuPPN84"
UU Perbendaharaan Indonesia (Indische Comptabiliteitswet)	<b>JIKA</b>	Token[i].Start="indicomp"
	<b>MAKA</b>	referensi="[ID_dokumen]->indcomp"

Berdasarkan aturan pada Tabel 3.10 di atas dapat dilihat bahwa setiap penemuan referensi dokumen yang sedang diproses ke dokumen lain dinyatakan dengan keterhubungan ID dokumen tersebut dengan ID dokumen lain menggunakan tanda panah '->'.  


### 3.16 Pembuatan Dokumen DOT

Setelah semua referensi yang dimiliki dokumen ditemukan maka proses selanjutnya adalah pembuatan dokumen DOT. Pembuatan dokumen ini dilakukan dengan cara menyatukan seluruh referensi yang ditemukan sebelumnya dalam satu berkas (*file*) dan menyimpannya dalam format *.dot*. Namun untuk memenuhi syarat pembentukan dokumen DOT, seluruh referensi tersebut diletakkan di antara tanda kurung kurawal dan di bagian awal dokumen ditambahkan kata kunci *digraph* diikuti nama graf. Selain itu dapat juga dilakukan penambahan atribut pada *node* dan *edge* dari graf untuk memperindah tampilan graf. Contoh dokumen DOT yang dihasilkan sistem dapat dilihat pada Lampiran 8.

Dokumen DOT yang dihasilkan berisi pendefinisian properti yang dimiliki suatu graf referensi. Dokumen tersebut hanya terbatas pada pendefinisian hubungan antar-*node* yang terdapat dalam graf. Untuk memvisualisasikan keterhubungan tersebut, maka *file* DOT dapat diproses lebih lanjut dengan

menggunakan aplikasi *Graphviz*. Hasil visualisasi graf yang didefinisikan dokumen DOT yang terdapat pada Lampiran 8 dapat dilihat pada Lampiran 17.

### 3.17 Evaluasi Sistem

Evaluasi sistem yang dilakukan pada penelitian ini bertujuan untuk mengukur keefektifan aturan-aturan yang digunakan dalam standarisasi dokumen. Keefektifan yang dimaksud adalah keefektifan dalam menemukan subbagian-subbagian serta frase-frase pengacuan yang telah diubah dalam bentuk elemen XML.

Proses pertama dalam evaluasi sistem adalah proses pembuatan dokumen kunci. Dokumen kunci dibuat dengan cara mengubah dokumen dalam format teks yang akan dievaluasi ke dalam format XML secara manual. Dokumen inilah yang dibandingkan dengan dokumen XML yang dihasilkan oleh sistem atau disebut dengan dokumen keluaran. Hasil perbandingan dinyatakan dalam nilai *recall*, *precision*, dan *F-measure* seperti yang telah dijelaskan pada subbab 2.6.

Proses evaluasi yang dilakukan dibagi menjadi enam jenis, yakni:

1. Evaluasi elemen bagian judul
2. Evaluasi elemen bagian pembukaan
3. Evaluasi elemen bagian batang tubuh
4. Evaluasi elemen bagian penutup
5. Evaluasi penemuan elemen referensi
6. Evaluasi keseluruhan elemen

Evaluasi jenis pertama sampai jenis keempat dilakukan dengan cara membandingkan setiap elemen pada dokumen keluaran dengan elemen pada dokumen kunci. Proses perbandingan tersebut menghasilkan nilai *possible*, *actual*, *correct* dan *partial* yang dibutuhkan dalam perhitungan nilai *recall* dan *precision*.

Nilai *possible* menyatakan jumlah elemen yang terdapat dalam dokumen kunci dan nilai *actual* menyatakan jumlah elemen yang terdapat dalam dokumen keluaran. Nilai *correct* diperoleh jika elemen yang terdapat dalam dokumen keluaran cocok dengan elemen yang terdapat dalam dokumen kunci. Sedangkan nilai *partial* diperoleh jika elemen yang terdapat dalam dokumen keluaran tidak cocok secara keseluruhan dengan elemen pada dokumen kunci namun hanya sebagian. Dari keempat nilai tersebut dilakukan penghitungan nilai *recall* dan

*precision*. Dan dari nilai *recall* dan *precision* yang dihasilkan dilakukan penghitungan nilai *F-measure*.

Evaluasi jenis kelima yaitu evaluasi penemuan elemen referensi pada prinsipnya sama dengan keempat jenis evaluasi sebelumnya. Akan tetapi, dalam evaluasi ini *tag* pembuka di mana elemen referensi tersebut berada juga ikut dibandingkan. Contohnya pada elemen ayat berikut:

```
<AYAT NO="6" ID="uu-11-2006.ps-4.ay-6">
Tata cara penetapan kawasan khusus sebagaimana dimaksud pada
<REF ID="uu-11-2006.ps-4.ay-1"/> dan <REF ID="uu-11-2006.ps-4.ay-
2"/> dilakukan sesuai dengan peraturan perundang-undangan
</AYAT>
```

Pada elemen ayat di atas terdapat dua elemen REF, yaitu:

- <REF ID="uu-11-2006.ps-4.ay-1"/>
- <REF ID="uu-11-2006.ps-4.ay-2"/>

Dengan demikian, proses evaluasi dilakukan dengan cara perbandingan:

- <AYAT NO="6" ID="uu-11-2006.ps-4.ay-6"><REF ID="uu-11-2006.ps-4.ay-1"/>
- <AYAT NO="6" ID="uu-11-2006.ps-4.ay-6"><REF ID="uu-11-2006.ps-4.ay-2"/>

Selain itu, pada evaluasi ini juga terdapat perbedaan pada jenis nilai yang dihitung. Jika pada keempat evaluasi sebelumnya nilai partial dihitung, maka pada evaluasi ini nilai tersebut tidak dihitung. Di sini nilai partial adalah nol. Sedangkan nilai correct diperoleh jika elemen REF pada dokumen keluaran beserta *tag* pembuka di mana elemen tersebut berada cocok dengan elemen REF dan *tag* pembuka pada dokumen kunci. Contohnya, jika elemen ayat di atas merupakan elemen pada dokumen kunci dan elemen pada dokumen keluarannya adalah sebagai berikut:

```
<AYAT NO="6" ID="uu-11-2006.ps-4.ay-6">
Tata cara penetapan kawasan khusus sebagaimana dimaksud pada
ayat (1) dan <REF ID="uu-11-2006.ps-4.ay-2"/> dilakukan sesuai
dengan peraturan perundang-undangan
</AYAT>
```

Hasil perbandingan elemen pada dokumen keluaran dengan elemen pada dokumen kunci di atas menghasilkan nilai *possible* adalah 2, nilai *actual* 1, dan nilai *correct* 1. Berdasarkan nilai *possible*, *actual*, dan *correct* tersebut, dilakukan penghitungan nilai *recall*, *precision*, dan *F-Measure* sebagai berikut:

$$\begin{aligned} \text{Recall} &= (\text{correct} + 0.5 * \text{partial}) / \text{possible} = (1 + 0) / 2 = 1 / 2 = 50\% \\ \text{Precision} &= (\text{correct} + 0.5 * \text{partial}) / \text{possible} = (1 + 0) / 1 = 1 / 1 = 100\% \\ \text{F-Measure} &= (\text{recall} * \text{precision}) / (0.5 * (\text{recall} + \text{precision})) \\ &= (50 * 100) / (0.5 * (50 + 100)) = 66.66\% \end{aligned}$$

Evaluasi jenis keenam, yaitu evaluasi keseluruhan elemen dilakukan pada seluruh elemen dalam dokumen mulai dari elemen pada bagian judul sampai elemen pada bagian penutup. Prinsip evaluasi ini sama dengan evaluasi jenis kedua sampai keempat. Tujuan dari evaluasi ini adalah untuk mengetahui keefektifan sistem dalam mengidentifikasi bagian-bagian serta subbagian-subbagian dalam dokumen secara keseluruhan.

