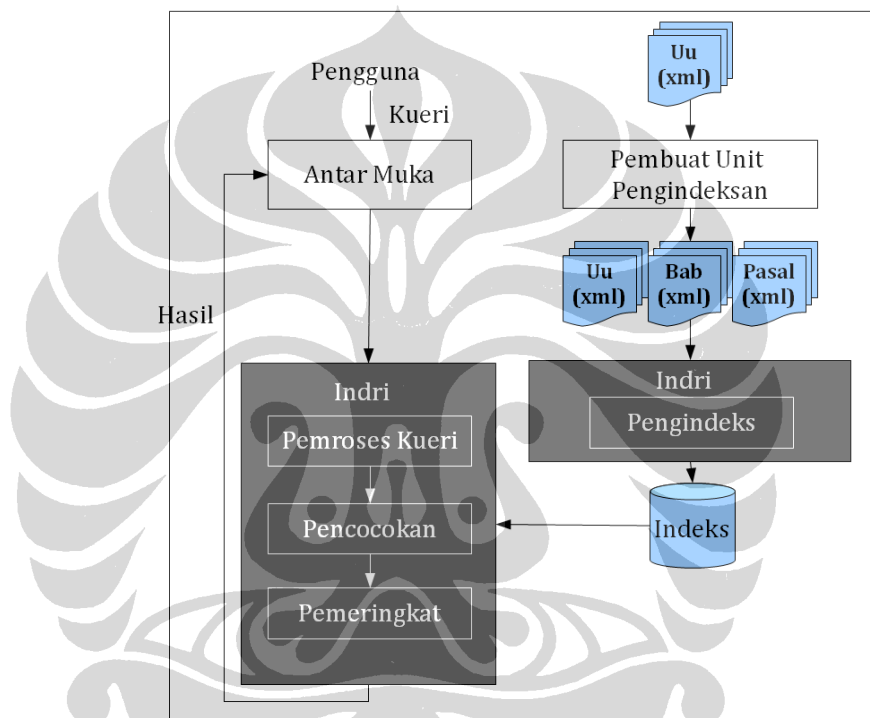


BAB 3 PERANCANGAN SISTEM

Bab ini menguraikan perancangan sistem yang dikembangkan meliputi arsitektur sistem, perancangan pembuat unit pengindeksan dan perancangan antarmuka.

3.1 Arsitektur Sistem



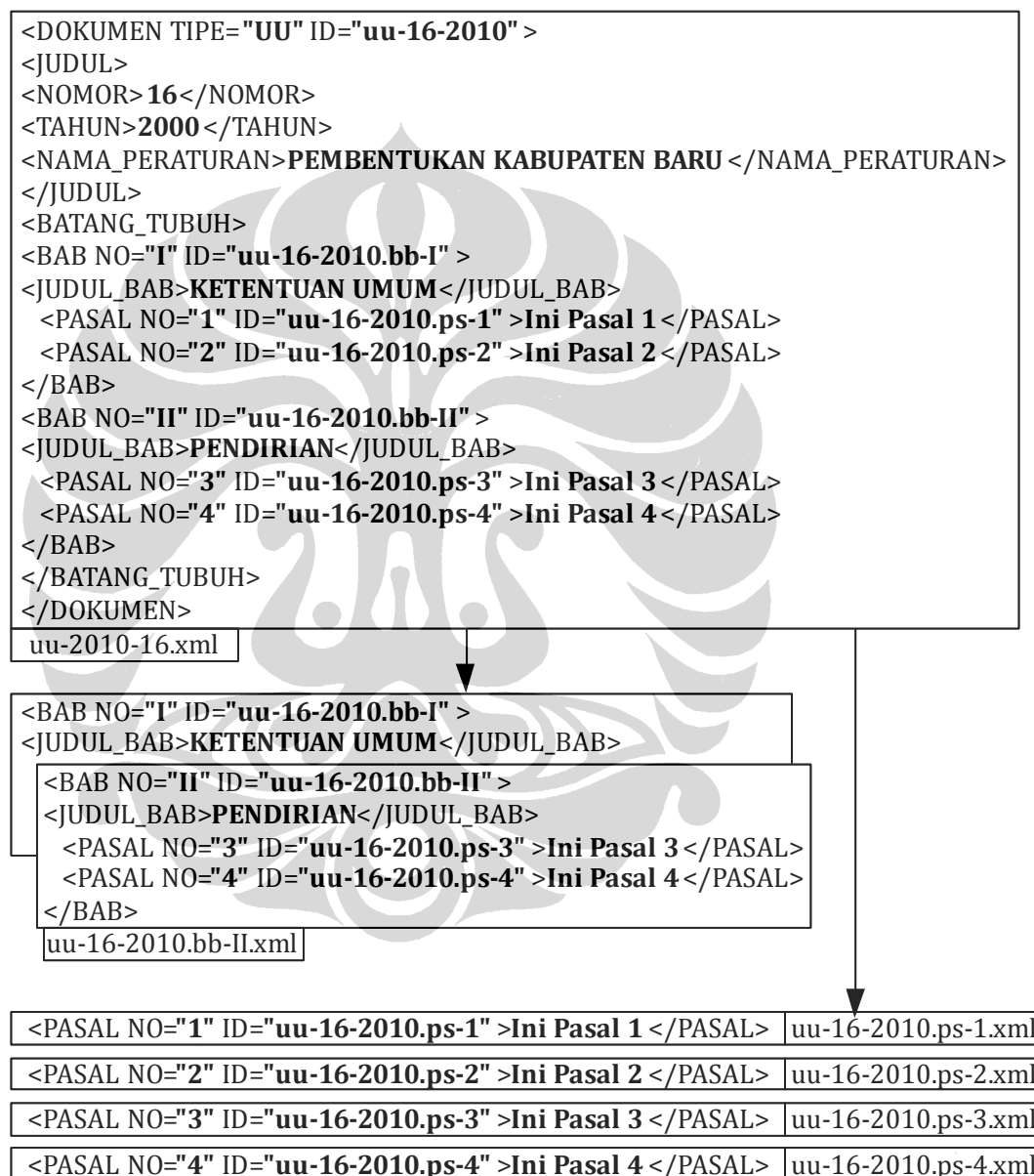
Gambar 3.1. Arsitektur sistem.

Berdasarkan Gambar 3.1, sistem yang dirancang terdiri dari empat bagian utama, yakni pembuat unit pengindeksan, pengindeks dokumen, fungsi perolehan, dan antarmuka. Pembuat unit pengindeksan berperan dalam mempersiapkan data yang akan digunakan. Pengindeks dokumen dan fungsi perolehan merupakan program inti sistem yang dikembangkan. Sementara antarmuka merupakan penghubung antara pengguna dan sistem. Pada bab ini, tidak dibahas secara rinci perancangan *Indri search engine*. *Indri search engine* dianggap sebagai sebuah *blackbox* yang diketahui masukan dan keluarannya saja. Konfigurasi penggunaan

Indri dibahas di bab berikutnya. Sementara penjelasan tentang model perolehan informasi yang digunakan oleh Indri telah dibahas di subbab 2.5.

3.2 Rancangan Pembuat Unit Pengindeksan

Sesuai dengan namanya, program ini berfungsi untuk membuat unit pengindeksan yakni elemen bab dan elemen pasal yang ada dalam undang-undang.



Gambar 3.2. Ilustrasi pembuatan unit bab dan unit pasal.

Untuk itu, penulis membuat program yang mampu membaca dokumen undang-undang XML, mengenali bagian bab dan bagian pasal dari undang-undang yang dibaca. Setelah dikenali, program akan membentuk sebuah berkas baru

berdasarkan elemen tersebut. Proses pengenalan bab dan pasal dari undang-undang dibahas pada subbab 3.1.1 dan 3.1.2.

3.2.1 Unit Pengindeksan Bab

Sebuah bab dalam undang-undang dengan format XML yang dibuat oleh Violina, dikenali dengan *tag* XML bernama bab. *Tag* ini memiliki *tag* pembuka dan *tag* penutup, yakni `<BAB>` dan `</BAB>`. *Tag* pembuka memiliki dua buah atribut, NO dan ID. *Value* dari atribut ID digunakan sebagai nama dari berkas yang memuat elemen bab tersebut. Selain memuat *tag* pembuka dan penutup bab, sebuah berkas baru juga memuat semua teks yang berada di antara kedua *tag* tersebut. Undang-undang yang tidak memiliki elemen bab akan diabaikan.

Masukan program unit pembuat unit pengindeksan bab adalah berkas berisi daftar undang-undang yang akan diproses. Setiap undang-undang akan dibuka dan dibaca isinya baris per baris. Pengenalan terhadap *tag* pembuka dan penutup bab dilakukan dengan mengecek isi setiap baris. *Tag* pembuka dikenali apabila baris yang dibaca memuat teks berikut.

```
<BAB NO="
```

Sementara *tag* penutup dikenali apabila suatu baris memuat teks berikut.

```
</BAB>
```

Pengambilan *value* atribut ID dilakukan dengan mengenali ciri atribut ID, yakni:

- (a) Didahului dengan: `<BAB NO="[IVX]*" ID="`
- (b) Diakhiri dengan: `">`

Semua karakter yang memenuhi syarat tersebut merupakan *value* atribut ID. *Value* atribut NO merupakan bilangan romawi yang menandakan nomor dari BAB. Pada poin (a), *value* atribut NO direpresentasikan dalam ekspresi reguler `"[IVX]*"` yang dapat diterjemahkan sebagai: "karakter I, V, atau X, diperbolehkan untuk muncul minimal 0 kali dan tidak ada batas maksimal". Ekspresi tersebut mewakili bilangan romawi mulai dari I hingga bilangan romawi terbesar yang dapat direpresentasikan dengan simbol I, V, atau X adalah XXXIX (39). Apabila menyertakan simbol V maka bilangan terbesarnya adalah XXXVIII (38). Belum

ditemukan undang-undang dengan penomoran bab hingga 38 sehingga tiga simbol tersebut cukup untuk mewakili semua penomoran bab undang-undang dalam koleksi.

Proses lain dalam pembuatan unit pengindeksan bab adalah menyalin semua teks yang berada di antara *tag* pembuka dan *tag* penutup bab. Proses ini cukup krusial mengingat penyertaan teks lain di luar kedua elemen tersebut akan membuat berkas bab yang dibentuk tidak valid karena tidak sama dengan aslinya. Salah satu solusi masalah tersebut yakni dengan menyertakan semacam *counter*. *Counter* diinisiasi dengan nilai 0. Apabila bertemu dengan *tag* pembuka bab, maka nilai *counter* bertambah 1. Apabila bertemu dengan *tag* penutup bab, nilai *counter* dikurangi 1. Penyalinan isi bab hanya boleh dilakukan apabila *counter* tidak bernilai 0. Proses ini diilustrasikan oleh Gambar 3.3.

```

$counter = 0 //inisiasi counter
<BAB NO="IV" ID="uu-11-2008.bb-IV"> $counter++; //counter = 1
...
...
</BAB> $counter--; //counter = 0

```

Gambar 3.3. Penggunaan *counter* untuk menjaga kevalidan isi elemen pasal.

3.2.2 Unit Pengindeksan Pasal

Ada dua variasi isi elemen pasal, yakni pasal inti undang-undang perubahan dan pasal biasa undang-undang non perubahan.

3.2.2.1 Pasal Inti Undang-undang Perubahan

Undang-undang perubahan adalah undang-undang yang berfungsi melakukan perubahan terhadap undang-undang lain. Salah satu jenis aksi pengubahan adalah merubah pasal undang-undang lain. Pasal baru yang mengalami perubahan itu dimuat dalam pasal-pasal undang perubahan. Pasal undang-undang perubahan yang memuat pasal undang-undang lain yang dirubahnya selanjutnya disebut pasal inti.

Pasal inti mempunyai ciri khas, yakni *value* dari atribut NO berupa angka romawi (I, II, III, ...), bukan angka arab (1, 2, 3, ...). Pasal inti inilah yang menjadi unit pengindeksan. Pasal yang dirubah tidak menjadi unit pengindeksan tersendiri, tetapi menjadi bagian dari unit pengindeksan pasal inti. Contoh pasal inti diperlihatkan pada Gambar 3.4.

```

<PASAL NO="I" ID="uu-1-1988.ps-I">
<REF ID="uu-20-1982.bb-III"/> tentang Ketentuan-ketentuan Pokok Pertahanan
Keamanan Negara Republik Indonesia (Lembaran Negara Tahun 1982 Nomor 51,
Tambahan Lembaran Negara Nomor 3234) diubah sehingga menjadi berbunyi
sebagai berikut:
  <PASAL NO="21" ID="uu-20-1982.ps-21">
    <AYAT NO="1" ID="uu-20-1982.ps-21.ay-1">
      Anggota Angkatan Bersenjata diperoleh secara sukarela dan wajib
      dari warga negara yang memenuhi persyaratan
    </AYAT>
    <AYAT NO="2" ID="uu-20-1982.ps-21.ay-2">
      Ketentuan-ketentuan tentang anggota Angkatan Bersenjata
      sebagaimana dimaksud <REF ID="uu-20-1982.ps-21.ay-1"/>
      diatur lebih lanjut dengan undang-undang
    </AYAT>
  </PASAL>
</PASAL>

```

Gambar 3.4. Contoh pasal inti undang-undang perubahan.

3.2.2.2 Pasal Biasa Undang-undang Non Perubahan

Pasal jenis ini merupakan pasal yang umum ditemukan. Cirinya yaitu *value* dari atribut NO berupa angka arab (1, 2, 3, ...) seperti ditunjukkan oleh Gambar 3.5. Pasal biasa beserta teks isinya inilah yang menjadi unit pengindeksan.

```

<PASAL NO="2" ID="uu-42-2008.ps-2">
Pemilu Presiden dan Wakil Presiden dilaksanakan
secara efektif dan efisien berdasarkan asas langsung, umum, bebas, rahasia, jujur,
dan adil
</PASAL>

```

Gambar 3.5. Pasal biasa undang-undang non perubahan.

Baik pasal inti maupun pasal biasa, sama-sama memiliki ciri umum seperti dibuka oleh teks:

```

<PASAL NO="

```

dan diakhiri oleh teks:

```
</PASAL>
```

Pengambilan *value* atribut ID pasal inti dilakukan dengan mengenali ciri atribut ID, yakni:

(a) Didahului dengan: <PASAL NO="I]+ " ID="

(b) Diakhiri dengan: ">

Pada poin (a), *value* atribut NO direpresentasikan dalam ekspresi reguler "[I]+" yang dapat diterjemahkan sebagai: "karakter I diperbolehkan untuk muncul minimal 1 kali dan tidak ada batas maksimal". Nomor terbesar pada pasal inti tidak pernah melebihi nilai yang diwakili bilangan romawi terbesar yang dapat direpresentasikan oleh karakter I, yakni III (3). Sementara itu, ciri dari atribut ID pada pasal biasa mempunyai kemiripan dengan pasal inti. Perbedaannya terletak pada *value* atribut NO yang direpresentasikan dalam ekspresi reguler sebagai "[0-9]*". Ekspresi tersebut cukup untuk merepresentasikan semua bilangan asli yang digunakan sebagai nomor pasal.

Penyalinan isi pasal dimulai sejak *tag* pembuka pasal inti atau pasal biasa undang-undang non perubahan. Untuk mencegah penyalinan diluar ketentuan itu, termasuk menganggap pasal yang dirubah pada pasal inti sebagai unit pengindeksan, digunakanlah *counter*. Prinsip *counter* ini sama dengan *counter* pada unit BAB, yakni diinisiasi dengan 0, bertambah 1 apabila bertemu dengan *tag* pembuka, dan berkurang 1 apabila bertemu dengan *tag* penutup. Proses tersebut diilustrasikan oleh Gambar 3.6.

```
$counter = 0 //inisiasi counter
<PASAL NO="I" ID="uu-11-2008.bb-I"> $counter++; //counter = 1
...
<PASAL NO="16" ID="uu-22-1999.ps-16"> $counter++; //counter = 2
...
</PASAL> $counter--; //counter = 1
...
</PASAL> $counter--; //counter = 0
```

Gambar 3.6. Penggunaan *counter* untuk mengetahui akhir dari elemen pasal.

3.3 Rancangan Konfigurasi Pengindeks Indri

Pengindeks Indri digunakan dalam merepresentasikan koleksi dokumen dalam struktur data yang lain sehingga proses pencarian menjadi lebih efisien. Program Pengindeks Indri merupakan program *built-in* dalam *Indri search engine*. Pengguna *Indri search engine* cukup mendefinisikan beberapa informasi yang akan digunakan dalam eksekusi program pengindeks. Informasi-informasi tersebut yakni *path* tempat koleksi dokumen, format koleksi dokumen, memori, *path* tempat menyimpan indeks yang dibuat, dan daftar *stopwords* yang digunakan. Semua informasi tersebut dimuat dalam sebuah berkas *parameter*.

Salah satu informasi yang didefinisikan dalam berkas *parameter* adalah *stopwords* atau daftar kata-kata yang umum ditemukan dalam koleksi dan tidak penting. Salah satu isu dalam perancangan sistem perolehan informasi untuk dokumen legal Indonesia adalah penggunaan *stopwords* khusus sebagaimana dibahas dalam subbab 1.1. Kata-kata seperti “menimbang”, “undang-undang”, “republik”, “indonesia”, “tentang”, “memutuskan”, termasuk dalam *stopwords* khusus tersebut karena kata-kata tersebut dapat dipastikan akan muncul dalam semua dokumen undang-undang. Namun demikian, dalam koleksi undang-undang berformat XML yang digunakan dalam penelitian ini, kata-kata tersebut tidak muncul karena direpresentasikan dalam sintaks-sintaks XML seperti ditunjukkan oleh Gambar 3.7.

<p>Penulisan bagian judul undang-undang:</p> <p>UNDANG-UNDANG REPUBLIK INDONESIA NOMOR 11 TAHUN 2008 TENTANG INFORMASI DAN TRANSAKSI ELEKTRONIK</p> <p>Penulisan bagian judul undang-undang dalam format XML:</p> <pre><JUDUL> <NOMOR>11</NOMOR> <TAHUN>2008</TAHUN> <NAMA_PERATURAN> INFORMASI DAN TRANSAKSI ELEKTRONIK </NAMA_PERATURAN> </JUDUL></pre>

Gambar 3.7. Contoh perubahan penulisan dalam undang-undang XML.

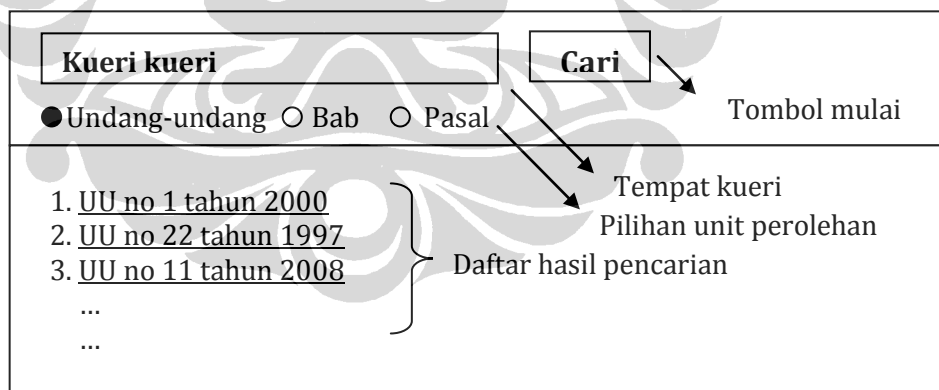
Adanya representasi dalam sintaks XML untuk kata-kata tersebut membuat *stopwords* khusus untuk dokumen legal tidak diperlukan lagi sehingga dalam penelitian ini *stopwords* yang digunakan adalah daftar kata-kata umum saja seperti “yang”, “dari”, dan “untuk”.

3.4 Rancangan Antarmuka


Sistem perolehan informasi yang dikembangkan penulis merupakan sistem yang berinteraksi dengan pengguna. Antarmuka yang dirancang memfasilitasi proses pencarian dan proses pengindeksan, tetapi tidak memfasilitasi proses pembuatan unit pengindeksan. Pada proses pencarian, penulis merancang antarmuka yang mencakup hal-hal berikut.

1. pemberian masukan kueri oleh pengguna.
2. pemilihan pencarian berdasarkan jenis indeks.
3. memulai aksi pencarian.
4. menampilkan daftar hasil pencarian.
5. menampilkan rincian dari daftar hasil pencarian.

Rancangan antarmuka proses pencarian ditunjukkan Gambar 3.8 dan Gambar 3.9.



Gambar 3.8. Rancangan antarmuka halaman pencarian.

Nama berkas: uu-11-2008.ps-13 Kueri: Sertifikasi Elektronik
Pasal 13 (1) Setiap Orang berhak menggunakan jasa Penyelenggara Sertifikasi Elektronik untuk pembuatan Tanda Tangan Elektronik (2) Penyelenggara Sertifikasi Elektronik harus memastikan keterkaitan suatu Tanda Tangan Elektronik dengan pemiliknya
 Rincian hasil pencarian

Gambar 3.9. Rancangan antarmuka rincian hasil pencarian.

Sementara itu, untuk proses pengindeksan, antarmuka dirancang untuk memfasilitasi pemilihan jenis dokumen yang akan diindeks dan memulai aksi pengindeksan. Rancangan antarmukanya ditunjukkan oleh Gambar 3.10.

Pilihan jenis dokumen yang akan diindeks: <input type="radio"/> Undang-undang <input checked="" type="radio"/> Bab <input type="radio"/> Pasal	<input type="button" value="Proses"/>
---	---------------------------------------

Gambar 3.10. Rancangan antarmuka halaman pengindeksan.

BAB 4 IMPLEMENTASI

Bagian ini menguraikan implementasi sistem perolehan informasi dengan domain dokumen undang-undang berformat XML. Implementasi sistem dimulai dengan membuat terlebih dahulu program untuk tahap *pre-processing*, yakni membuat unit pengindeksan bab dan pasal yang ditulis dengan bahasa pemrograman Perl. Kemudian konfigurasi terhadap *Indri search engine* yang digunakan dalam proses pengindeksan dan perolehan informasi. Langkah selanjutnya adalah implementasi antarmuka berbasis web menggunakan *script* PHP, HTML, dan CSS.

4.1 Implementasi Pembuat Unit Pengindeksan

Subbab ini membahas implementasi program untuk membuat unit pengindeksan bab dan pasal. Subbab 4.1.1 dan subbab 4.1.2 berturut-turut membahas implementasi pembuatan unit pengindeksan bab dan pasal.

4.1.1 Implementasi Unit Pengindeksan Bab

Subbab ini membahas implementasi program untuk membuat unit pengindeksan bab. Langkah-langkah implementasinya sebagai berikut.

1. Buka berkas berisi daftar undang-undang beserta *path*-nya.
2. Buka sebuah berkas undang-undang yang ada dalam daftar.
3. Baca baris per baris berkas undang-undang yang dibuka..
4. Apabila bertemu dengan baris berisi *tag* pembuka bab, simpan ID dari *tag* tersebut, dan salin isi baris itu ke suatu variabel string SS.
5. Salin isi bab ke variabel string SS.
6. Apabila bertemu dengan baris berisi *tag* penutup bab, salin isi baris itu ke variabel SS, buat berkas baru dengan nama ID bab, tulis isi variabel SS ke berkas baru tersebut.
7. Apabila masih ada undang-undang dalam daftar yang belum diproses, kembali ke langkah 2.

Source code untuk implementasi pembuatan unit pengindeksan bab menggunakan bahasa pemrograman Perl dapat dilihat pada Gambar 4.1.

```

#!/usr/bin/perl
sub process
{
    my ($filename) = @_;
    open IN, $filename;

    foreach $line (<IN>)
    {
        if ($line =~ /#include (.*)/)
        {
            $tag = 0; #inisiasi counter
            process $1;
            next;
        }
        #mengenal tag pembuka bab
        elsif ( $line =~ /\<BAB NO="[IVX]" ID="(.)"/)
        {
            $tag++; #counter = 1
            $dokid = $1;
            #dokid dokumen adalah value dari atribut ID
            $sisi = $line;
        }
        #mengenal tag penutup bab
        elsif ( $line =~ /\<\/BAB\>/)
        {
            $tag--; #counter = 0
            $sisi = $sisi.$line;
            #buat file
            open(OUT, ">./test-bab/$dokid".xml");
            print OUT $sisi; #tulis isi ke file
            close(OUT);
            $sisi = ""; #kosongkan isi
        }
        elsif ($tag != 0)
        {
            #menyalin isi bab selain tag
            $sisi = $sisi.$line;
        }
    }
    close IN;
}
#membuka berkas berisi daftar nama undang-undang yang akan diproses
open(IN, "uu-xml-list.txt");
@filenames = <IN>; #menyalin daftar nama file ke array
foreach $file (@filenames){ process $file; }

```

Gambar 4.1. Kode implementasi pembuatan unit pengindeksan bab.

4.1.2 Implementasi Unit Pengindeksan Pasal

Subbab ini membahas implementasi program untuk membuat unit pengindeksan pasal, baik pasal inti undang-undang perubahan maupun pasal biasa undang-undang non perubahan. Langkah-langkah implementasinya sebagai berikut.

1. Buka berkas berisi daftar undang-undang beserta *path*-nya.
2. Buka sebuah berkas undang-undang yang ada dalam daftar.
3. Baca baris per baris berkas yang dibuka.
4. Apabila bertemu dengan baris berisi *tag* pembuka pasal, indentifikasi apakah pasal tersebut merupakan salah satu dari tiga kemungkinan berikut:
 - (a) pasal inti undang-undang perubahan
 - (b) pasal biasa undang-undang non perubahan
 - (c) pasal yang dirubah dalam undang-undang perubahan

Apabila termasuk ke poin (a) atau poin (b): simpan ID dari *tag* tersebut, dan salin isi baris itu ke suatu variabel string SS. Apabila termasuk poin (c) ID *tag* tidak perlu disimpan, salin saja isinya ke variabel SS.

5. Salin hanya isi pasal ke variabel string SS.
6. Apabila bertemu dengan baris berisi *tag* penutup pasal, indentifikasi apakah *tag* tersebut merupakan *tag* penutup salah satu jenis pasal pada langkah 3. Apabila termasuk ke pasal inti atau pasal biasa: salin isi baris itu ke variabel SS, buat berkas baru dengan nama ID pasal, tulis isi variabel SS ke berkas baru tersebut. Apabila bukan, salin saja baris itu ke variabel SS.
7. Apabila masih ada undang-undang dalam daftar yang belum diproses, kembali ke langkah 2.

Source code untuk implementasi pembuatan unit pengindeksan pasal menggunakan bahasa pemrograman Perl dapat dilihat pada Gambar 4.2.

```
#!/usr/bin/perl
sub process
{
    my ($filename) = @_;
    $uu_perubahan = 0; $tag = 0;
    open IN, $filename;

    foreach $line (<IN>)
    {
        if ($line =~ /#include (.*)/){
            process $1;
            next;
        }
        #mengenal tag pembuka pasal inti, uu-perubahan
        elsif ( $line =~ /\<PASAL NO="[1]"* ID="(.)"/ ) {
            $tag++;
            $dokid = $1;      #dokid adalah value atribut id tag pasal
            $uu_perubahan = 1;
            $isi = $line;
        }
        #mengenal tag pembuka pasal, uu-non-perubahan
        elsif ( $line =~ /\<PASAL NO="[0-9]"* ID="(.)"/ and $uu_perubahan == 0 ) {
            $dokid = $1;      #dokid adalah value atribut id tag pasal
            $tag++; $isi = $line;
        }
        #mengenal pasal yang dirubah dalam uu-perubahan
        elsif ( $line =~ /\<PASAL NO="[0-9]"* ID="(.)"/
            and $uu_perubahan == 1 ) {
            $tag++;
            $isi = $isi.$line;
        }
        #mengenal tag penutup pasal
        elsif ( $line =~ /\<\/PASAL\>/){
            $tag--;          #tag dekremen
            $isi = $isi.$line;
            if ($tag == 0) {
                #buka file baru, namanya dari id pasal
                open(OUT, ">./test-pasal/$dokid". ".xml");
                print OUT $isi;      #tulisi isi ke file
                close(OUT);
                $isi = "";          #kosongkan isi
            }
        }
        elsif ($tag != 0)
        {
            $isi = $isi.$line;      #salin isi dari pasal
        }
    }
    close IN;
}
open(IN, "uu-xml-list.txt");      #membuka berkas daftar nama uu
@filenames = <IN>;              #menyalin daftar nama file ke array
foreach $file (@filenames){ process $file;}
```

Gambar 4.2. Kode implementasi pembuatan unit pengindeksan pasal.

4.2 Konfigurasi Pengindeks Indri Search Engine

Penelitian ini menggunakan Indri *search engine* sebagai alat bantu dalam proses mengindeks dokumen. Untuk menyesuaikan dengan kebutuhan perlu dilakukan konfigurasi program pengindeks Indri, yakni berkas *buildindex.exe*. Konfigurasi dilakukan dengan membuat sebuah *parameter file* yang disertakan saat eksekusi program pengindeks Indri. *Parameter file* yang dibuat digunakan untuk mendefinisikan beberapa hal berikut [OGI06]:

- (a) *path* ke korpus atau alamat tempat koleksi dokumen yang akan diindeks.
- (b) format dari koleksi dokumen.
- (c) *path* atau alamat untuk menyimpan hasil proses pengindeksan.
- (d) memori komputer yang digunakan saat pengindeksan.
- (e) daftar *stopwords* (opsional).

```

<parameters>
  <corpus>
    <!--relative path tempat koleksi dokumen berada-->
    <path>./test-uu</path>
    <!--format koleksi dokumen-->
    <class>xml</class>
  </corpus>
  <!--memory yang digunakan saat pengindeksan-->
  <memory>256m</memory>
  <!--relative path untuk menyimpan indeks-->
  <index>./index-uu</index>
  <!--daftar kata yang menjadi stopwords-->
  <stopper>
    <word>di</word>      <word>ke</word>      <word>dari</word>
    <word>dengan</word><word>dan</word>      <word>dalam</word>
    <word>tentang</word><word>lain</word>      <word>yang</word>
    <word>untuk</word> <word>pada</word>      <word>ini</word>
    <word>itu</word>    <word>telah</word>    <word>dimaksud</word>
    <word>adalah</word> <word>ialah</word>
    <word>merupakan</word><word>atas</word> <word>oleh</word>
    <word>paling</word> <word>secara</word> <word>serta</word>
    <word>atau</word>   <word>kali</word>   <word>saat</word>
    <word>perlu</word> <word>terdiri</word> <word>kepada</word>
    <word>karena</word> <word>bagi</word>   <word>bahwa</word>
  </stopper>
</parameters>

```

Gambar 4.3. Contoh *parameter file* pengindeksan undang-undang.

Potongan kode pada Gambar 4.3 menjelaskan 5 hal, yakni *path* tempat koleksi dokumen yang akan diindeks disimpan, format koleksi dokumen, memori, *path* tempat menyimpan indeks yang akan dibuat, dan daftar *stopwords* yang

digunakan. Alamat tempat menyimpan koleksi dokumen didefinisikan dalam `<path>./test-uu</path>`. Sementara alamat untuk menyimpan indeks disebutkan dalam `<index>./index-uu</index>`.

Konfigurasi berkas parameter untuk unit bab dan unit pasal tidak berbeda dengan unit undang-undang. Hanya perlu sedikit modifikasi pada alamat tempat koleksi dokumen disimpan serta alamat untuk menyimpan indeks yang dibuat. Contoh alamat tempat menyimpan koleksi dokumen unit bab adalah `<path>./test-bab</path>`. Tempat menyimpan indeks bab yang telah dibuat adalah `<index>./index-bab</index>`.

4.3 Konfigurasi Pemroses Kueri Indri

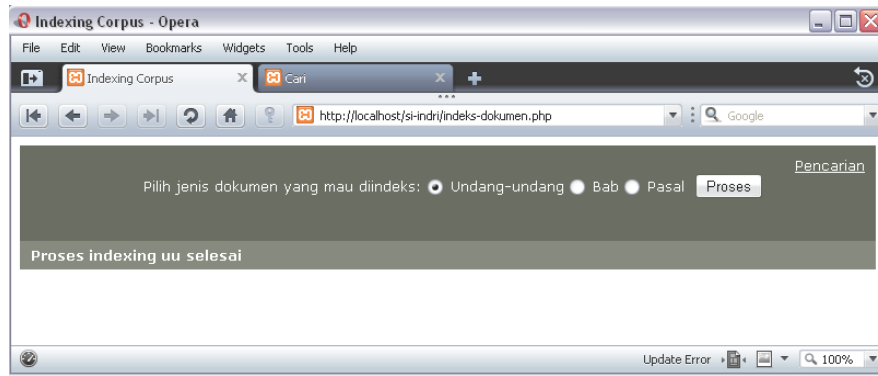
Konfigurasi program pemroses kueri (fungsi pencarian) mendefinisikan alamat tempat menyimpan indeks, jumlah hasil pencarian yang ingin ditampilkan, serta memori yang digunakan. Pada eksekusi program pemroses kueri Indri, yakni *file runquery.exe*, disertakan kueri yang akan dieksekusi. Contoh kodenya sebagai berikut.

```
runquery.exe -memory="256M" -count="10" -index="./index-pasal"
-query="presiden"
```

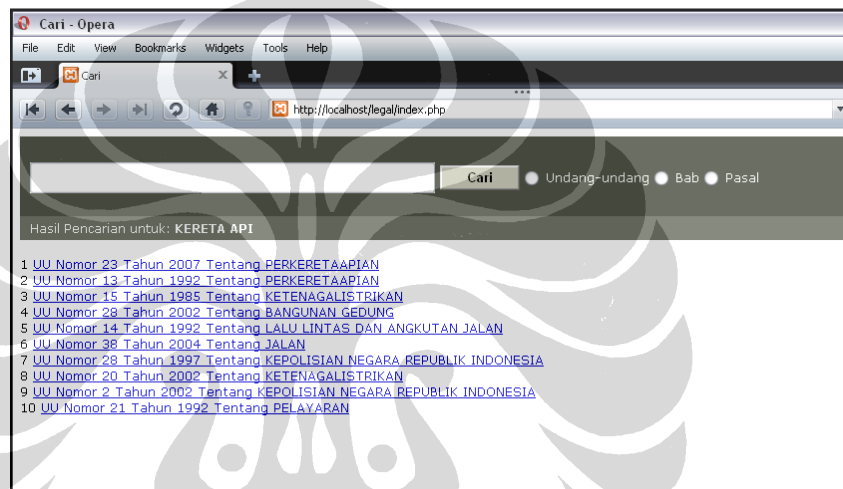
Parameter `-count` merepresentasikan jumlah hasil pencarian yang ditampilkan, `-index` berisi alamat tempat indeks disimpan, `-query` nilainya merupakan kueri perolehan informasi. Penulis menyertakan potongan kode tersebut dalam antarmuka untuk proses pencarian.

4.4 Implementasi Antarmuka

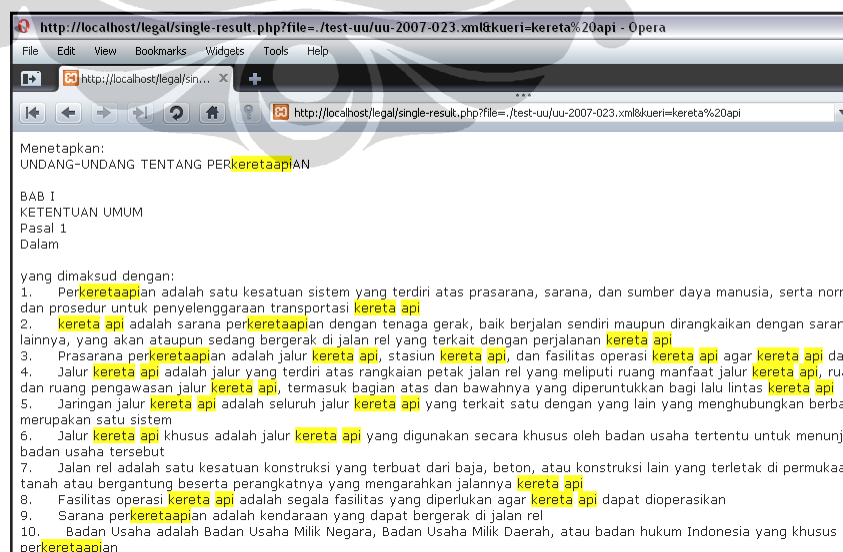
Sistem berbasis web ini dikembangkan dengan menggunakan bahasa HTML, PHP, dan CSS. Gambar 4.4, Gambar 4.5, dan Gambar 4.6 berturut-turut merupakan tampilan antarmuka untuk proses pengindeksan, proses pencarian, dan tampilan rincian hasil pencarian.



Gambar 4.4. Antarmuka halaman pengindeksan dokumen.



Gambar 4.5. Antarmuka halaman pencarian dan tampilan hasil pencarian.



Gambar 4.6. Antarmuka rincian hasil pencarian.

BAB 5 UJICоба DAN ANALISIS HASIL

Pada bab ini dijelaskan tentang tahap ujicoba dan analisis hasil ujicoba. Pembahasan meliputi koleksi data yang digunakan, skenario ujicoba, lingkungan ujicoba, hasil ujicoba, dan analisis hasil ujicoba.

5.1 Koleksi Data

5.1.1 Koleksi Dokumen

Penelitian ini menggunakan 488 berkas teks undang-undang berformat XML yang diundangkan dari tahun 1983 hingga tahun 2009. Undang-undang tersebut diperoleh dengan bantuan Hukum Online [HUK09]. Hukum Online memberikan koleksi undang-undang Republik Indonesia yang terbit sejak tahun 1950. Akan tetapi, undang-undang sebelum tahun 1983 memiliki masalah saat diproses menjadi dokumen XML sehingga dokumen yang dapat digunakan adalah undang-undang yang terbit mulai tahun 1984 hingga 2009. Setelah melalui tahap *pre-processing*, karakteristik data yang digunakan seperti ditunjukkan oleh Tabel 5.1.

Tabel 5.1. Karakteristik koleksi data.

Unit Pengindeksan	Jumlah Berkas	Ukuran Dokumen	Ukuran Total	Ukuran Rata-rata	Jumlah Kata Unik	Total Kata
Undang-undang	488	4 – 287 KB	15,3 MB	32 KB	14.480	1.329.179
Bab	3.102	1 – 153 KB	11,8 MB	4 KB	12.295	1.015.265
Pasal	15.646	1 – 45 KB	12 MB	1 KB	13.571	1.070.750

Ada 488 berkas undang-undang, 3.102 bab, dan 15.646 pasal yang kesemuanya berformat XML. Ukuran total data yang digunakan yakni 15,3 MB untuk koleksi undang-undang, 11,8 MB untuk bab, dan 12 MB untuk pasal. Ukuran dokumen pun bervariasi. Berkas undang-undang memiliki ukuran dokumen mulai dari 4 sampai 287 KB. Sementara untuk berkas bab dan pasal, ukurannya berturut-turut 1-153 KB dan 1-45 KB. Pasal-pasal yang dengan ukuran besar tersebut biasanya merupakan pasal inti undang-undang perubahan atau pasal yang menjelaskan ketentuan umum pada undang-undang non perubahan.

Jumlah kata unik pada unit undang-undang mencapai 14.480 kata, pada unit bab ada 12.295 kata, dan pada unit pasal ada 13.571 kata. Sementara itu, total kata yang digunakan pada unit undang-undang berjumlah 1.329.179, pada unit bab berjumlah 1.015.265, sedangkan unit pasal memiliki 1.070.750 kata. Baik jumlah kata unik maupun jumlah total kata, unit pengindeksan undang-undang memiliki nilai yang lebih tinggi ketimbang dua unit lainnya. Ini merupakan hal yang wajar, mengingat selain memiliki elemen bab dan pasal, unit undang-undang masih memiliki beberapa elemen lain, seperti judul, pembukaan, dan penutup. Jumlah kata unik dan total kata pada unit bab yang lebih rendah ketimbang unit pasal terjadi karena ada sebagian undang-undang yang hanya memiliki unit pasal, tetapi tidak memiliki bab.

5.1.2 Kueri

Kueri yang digunakan dalam penelitian berjumlah 50 buah. Penulis membuat sendiri kueri-kueri tersebut. Penulis membuat kueri dengan membaca sejumlah undang-undang, memilih topik yang menarik, kemudian mengambil kata kunci topik yang menarik tersebut untuk digunakan sebagai kueri. Kueri yang digunakan dalam pengujian disertakan sebagai lampiran laporan ini. Dalam pengujian, kueri tersebut tidak melalui proses *stemming* maupun penghilangan *stopwords*. Kueri yang digunakan bertipe boolean dengan operator AND. Operator tersebut tidak disertakan saat penulisan kueri karena Indri secara otomatis menambahkannya. Suatu dokumen hasil pencarian dinilai relevan apabila membahas topik dalam kueri.

5.2 Skenario Ujicoba

Penulis melakukan ujicoba sistem perolehan informasi dokumen legal dengan tahap sebagai berikut:

1. Mengindeks setiap jenis unit pengindeksan.
2. Melakukan perolehan informasi (pencarian) terhadap setiap jenis indeks menggunakan kueri yang sama.
3. Mengambil 10 dokumen teratas setiap hasil pencarian setiap kueri.

4. Melakukan penilaian (relevan atau tidak) terhadap 10 dokumen teratas untuk setiap kueri. Secara total, penulis menilai 1500 dokumen hasil pencarian.
5. Melakukan evaluasi hasil penilaian menggunakan metode MAP, P@5, dan P@10.

5.3 Lingkungan Ujicoba

Program berbasis web ini menggunakan Apache 2.2 sebagai server yang terintegrasi dalam perangkat lunak XAMPP [XAM08]. Sistem operasi yang dipergunakan adalah Windows XP Professional. Penulis melakukan implementasi pada komputer dengan spesifikasi prosesor Intel Pentium Dual CPU @ 2.00 GHz, 512 MB RAM, dan 80 GB *harddisk space*.

5.4 Hasil Ujicoba

Pada subbab ini dijelaskan hasil ujicoba sistem perolehan informasi dokumen legal yang penulis kembangkan. Hasil ujicoba yang dibahas merupakan nilai kinerja sistem yang diperoleh dengan menggunakan metode *mean average precision*, *precision at 5*, dan *precision at 10*. Hasil tersebut ditunjukkan oleh Tabel 5.2.

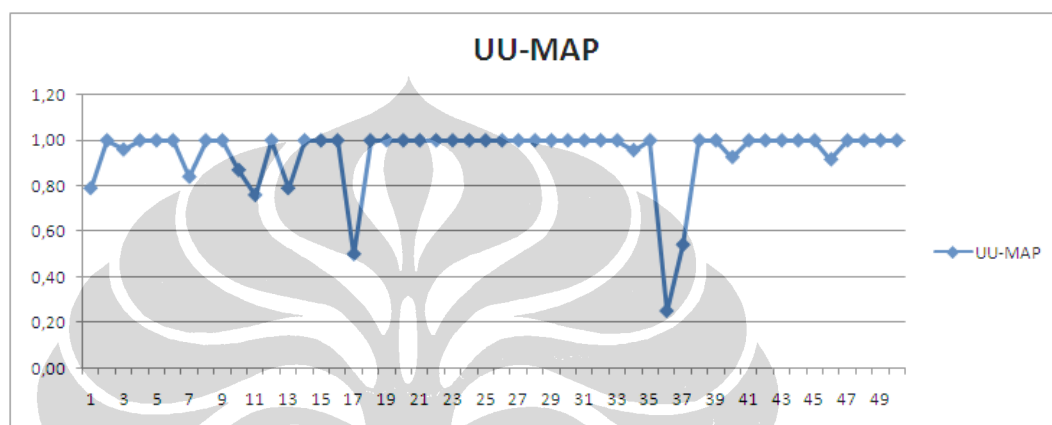
Tabel 5.2. Hasil ujicoba.

Unit Perolehan Informasi	MAP	P@5	P@10
Undang-undang	0,94	0,44	0,25
Bab	0,68	0,47	0,33
Pasal	0,72	0,52	0,43

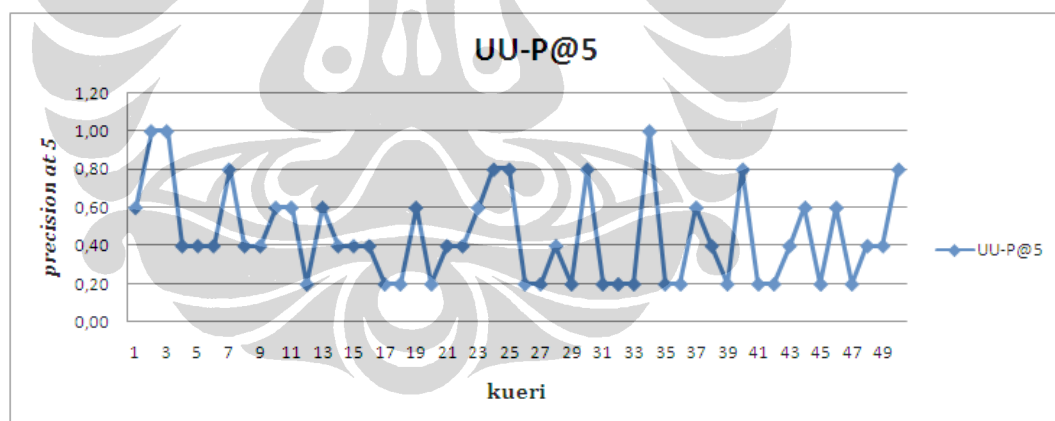
Tabel tersebut merupakan rata-rata dari presisi rata-rata, *precision at 5*, dan *precision at 10* dari 50 buah kueri. Undang-undang memiliki MAP tertinggi dibandingkan dua jenis indeks lainnya, tetapi pada pengukuran jenis lain undang-undang memberikan hasil yang paling rendah. Unit pasal justru memperoleh hasil lebih baik dibanding unit undang-undang dan unit bab dalam pengukuran *precision at 5* dan *precision at 10*. Dua pengukuran tersebut memberikan urutan peringkat yang sama untuk tiga jenis indeks.

5.4.1 Hasil Unit Undang-undang

Nilai MAP unit undang-undang yang hampir mendekati 1 menunjukkan bahwa unit undang-undang sering mendapatkan presisi rata-rata 1 pada 50 kueri uji coba. Hal tersebut dapat dilihat pada Gambar 5.1. Presisi rata-rata 1 dapat muncul akibat dokumen undang-undang yang relevan sering ditemukan pada peringkat pertama. Dokumen-dokumen relevan yang lain berada di belakangnya dan kemungkinan tanpa diselingi oleh dokumen non relevan.

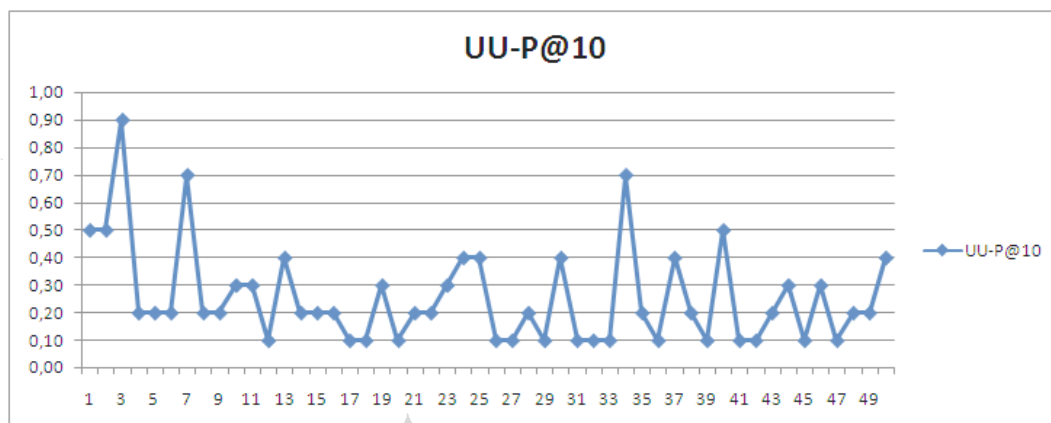


Gambar 5.2. MAP pada unit undang-undang.



Gambar 5.1. Precision at 5 unit undang-undang.

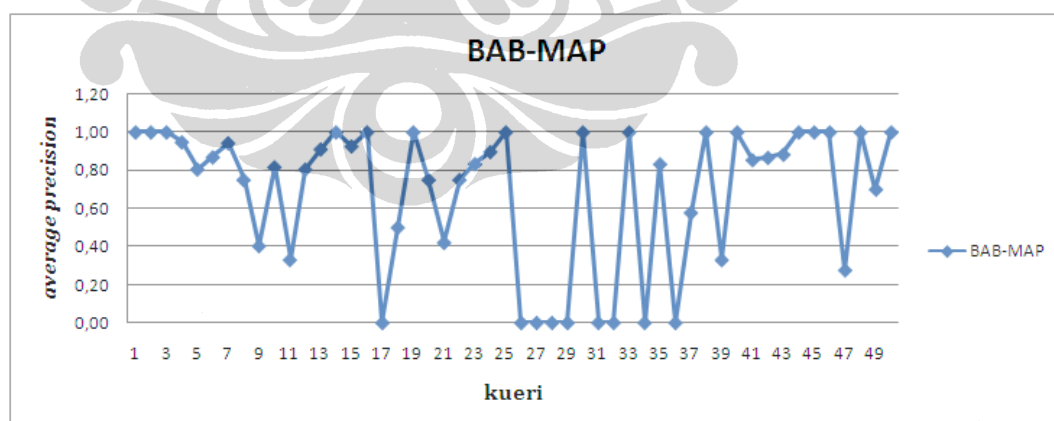
Gambar 5.2 menunjukkan hasil *precision at 5* pada 50 kueri untuk unit undang-undang. *Precision at 5* tersebut adalah 0,44. Hasil tersebut dapat diinterpretasikan bahwa dalam 5 dokumen teratas hasil pencarian akan ditemukan 2,2 dokumen relevan. Gambar itu juga menunjukkan bahwa untuk seluruh kueri, selalu ada dokumen undang-undang yang relevan. Sementara itu, hasil *precision at 10* pada 50 kueri unit undang-undang ditunjukkan oleh Gambar 5.3.



Gambar 5.3. *Precision at 10* unit undang-undang.

Nilai *precision at 10* unit undang-undang adalah 0,25. Nilai tersebut dapat diinterpretasikan bahwa dalam 10 buah dokumen teratas akan ditemukan 2,5 dokumen undang-undang yang relevan. Kecenderungan itu dapat dilihat pada Gambar 5.3. Hasil interpretasi *precision at 5* dan *precision at 10* unit undang-undang memiliki kesamaan dalam hal jumlah dokumen relevan yang ditemukan. Berdasarkan hal tersebut, dapat ditarik interpretasi lain yakni dokumen relevan pasti dapat ditemukan pada 5 dokumen teratas.

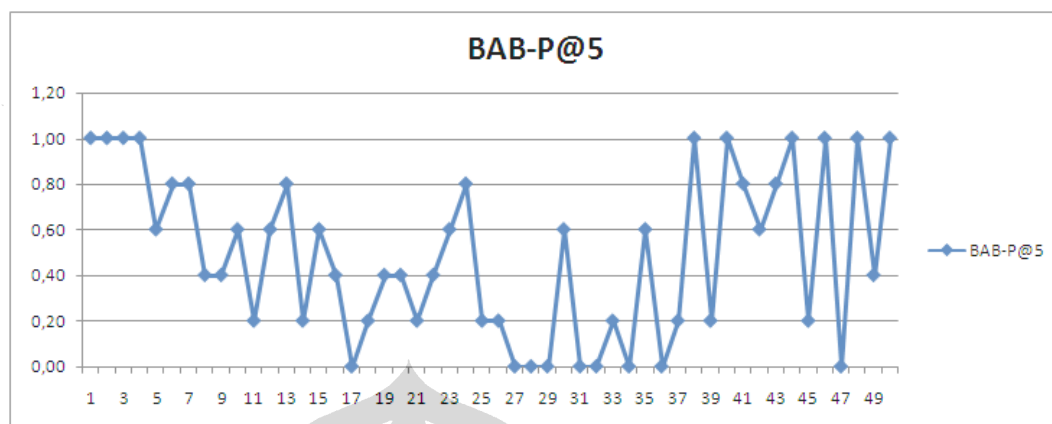
5.4.2 Hasil Unit Bab



Gambar 5.4. MAP unit bab.

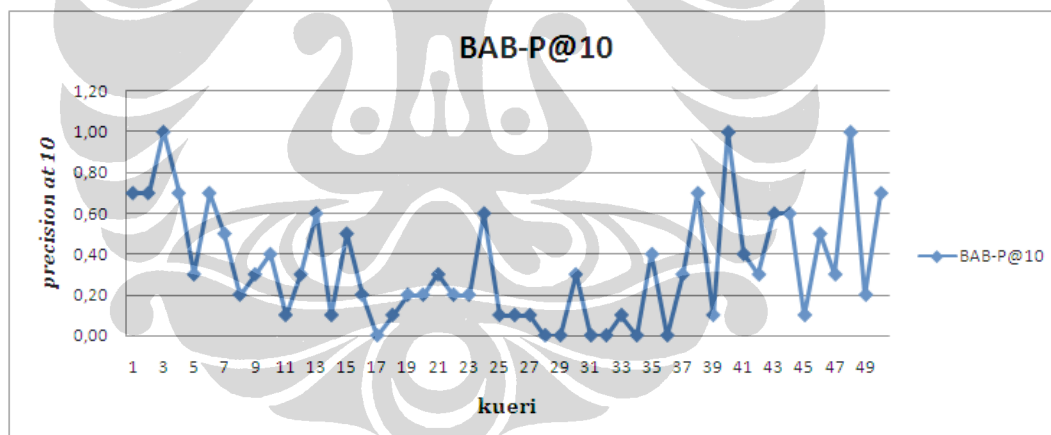
Unit bab mempunyai nilai MAP sebesar 0,68. Gambar 5.4 menunjukkan nilai *mean average precision* untuk unit bab pada 50 kueri. Pada gambar tersebut ada

cukup banyak kueri yang nilainya 1, tetapi ada sebagian yang lain yang nilainya 0. Lebih dari separuh ujicoba mempunyai MAP di atas 0,5.



Gambar 5.5. *Precision at 5* unit bab.

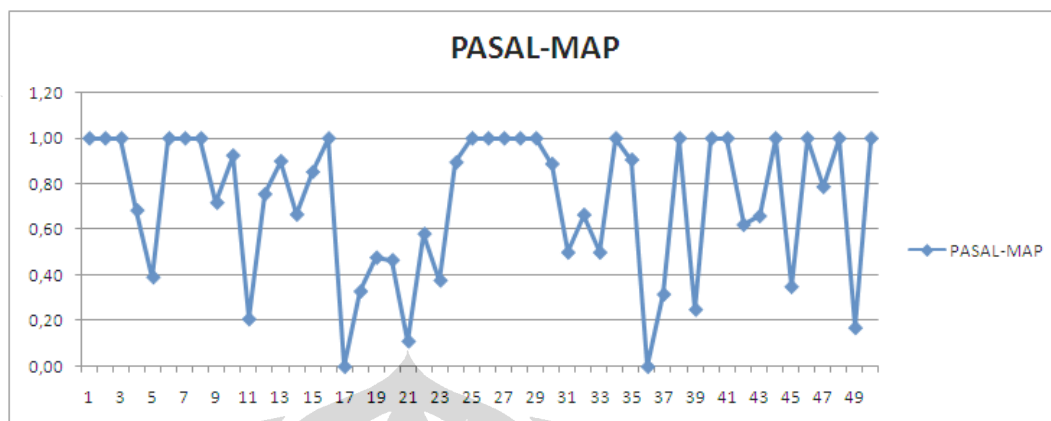
Nilai *precision at 5* dari unit perolehan informasi bab adalah 0,47. Jumlah kueri yang nilainya satu dan yang nilainya nol cukup berimbang. Sisanya berada di antara dua nilai tersebut. Interpretasi dari nilai *precision at 5* sebesar 0,47 yakni dalam 5 dokumen teratas terdapat 2,35 dokumen yang relevan.



Gambar 5.6. *Precision at 10* unit bab.

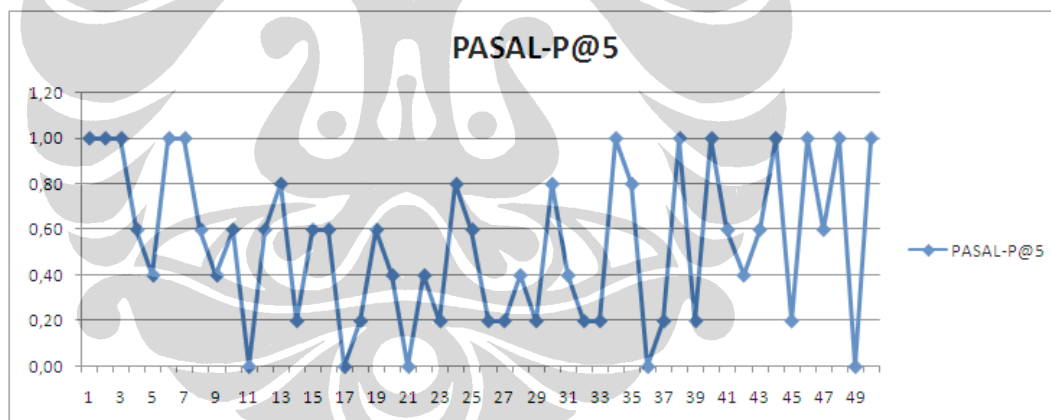
Sementara itu, nilai *precision at 10* unit bab yang bernilai 0,33 dapat diinterpretasikan bahwa 3,3 dari 10 dokumen teratas merupakan dokumen yang relevan. Jumlah tersebut 1 dokumen lebih banyak dibandingkan *precision at 5*. Secara intuitif dua hasil penilaian tersebut dapat menunjukkan bahwa 2 dari 3 dokumen relevan dalam 10 dokumen teratas, berada di peringkat 1 sampai 5.

5.4.3 Hasil Unit Pasal



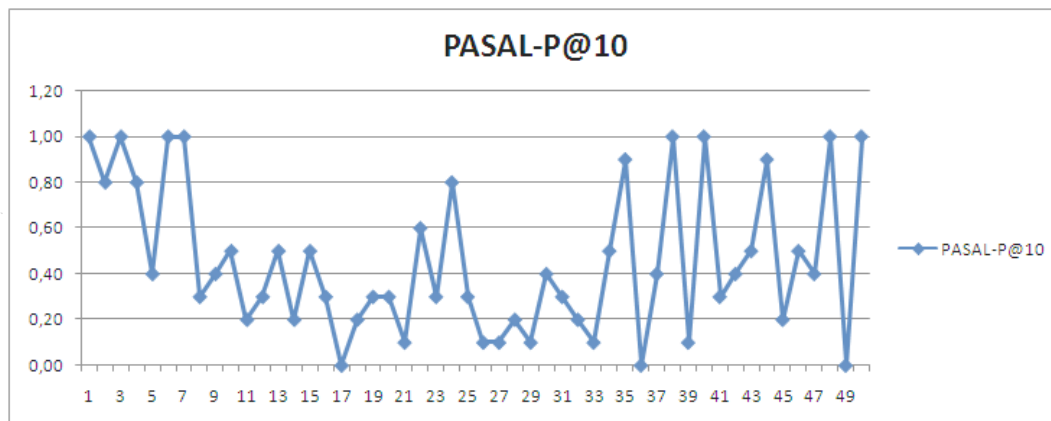
Gambar 5.7. MAP pada unit pasal.

Unit perolehan informasi pasal mempunyai MAP 0,72. Gambaran keseluruhan nilai rata-rata pada 50 kueri dapat dilihat pada Gambar 5.7. Kueri yang mempunyai nilai rata-rata 1 terlihat cukup mendominasi Gambar 5.7.



Gambar 5.8. *Precision at 5* pada unit pasal.

Untuk hasil *precision at 5* dan *precision at 10*, unit pasal memiliki nilai yang lebih tinggi dibandingkan dua unit lainnya. *Precision at 5* yang bernilai 0,52 menunjukkan bahwa lebih dari separuh, sekitar 3 dokumen relevan berada di deretan 5 dokumen teratas.

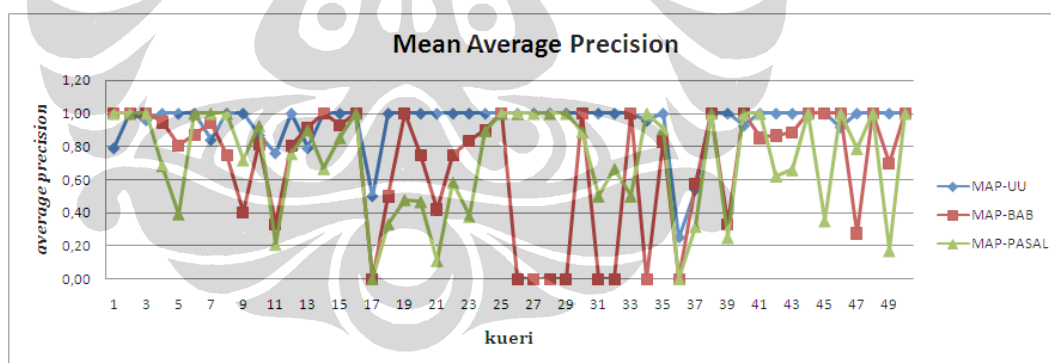


Gambar 5.9. Precision at 10 unit pasal.

Gambar 5.9 menunjukkan hasil *precision at 10* unit perolehan informasi pasal. *Precision at 10* yang bernilai 0,43 dapat diinterpretasikan bahwa 4,3 dari 10 dokumen teratas merupakan dokumen relevan.

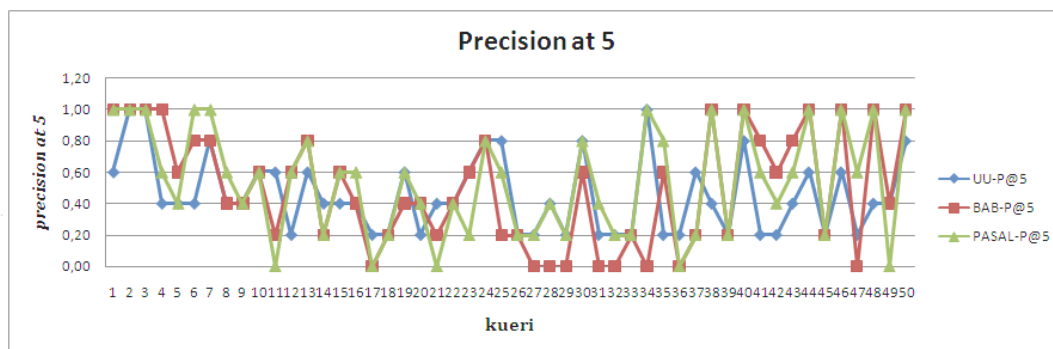
5.4.4 Perbandingan Unit Undang-undang, Unit Bab, dan Unit Pasal

Perbandingan hasil pada tiga jenis unit perolehan informasi dengan metode MAP, *precision at 5*, dan *precision at 10* dapat dilihat pada Gambar 5.10, Gambar 5.11, dan Gambar 5.12.



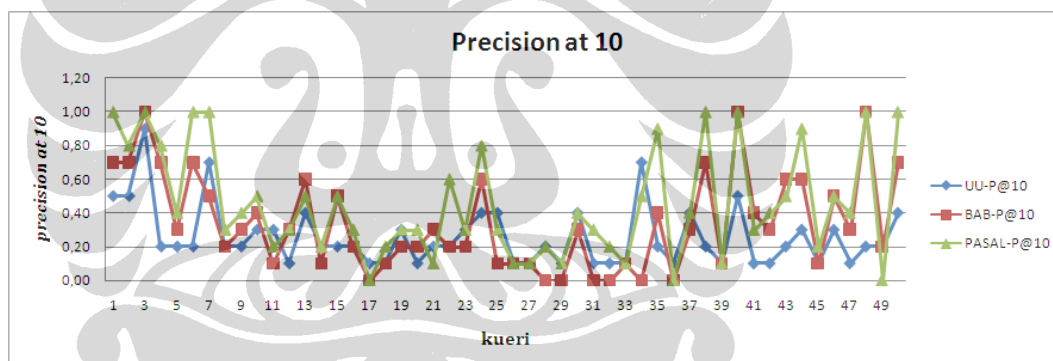
Gambar 5.10. MAP pada 3 unit perolehan informasi.

Berdasarkan Gambar 5.10, unit undang-undang terlihat paling sering mendapatkan nilai 1, sementara unit bab paling sering mendapatkan nilai 0. Pada kueri 26-29 dimana unit bab mendapat nilai 0, dua unit lainnya justru mendapatkan nilai 1.



Gambar 5.11. *Precision at 5* pada 3 unit perolehan informasi.

Precision at 5 pada 3 unit perolehan informasi mempunyai kecenderungan yang sama. Apabila ada satu unit yang memiliki nilai rendah pada sebuah kueri, unit yang lain juga cenderung memiliki nilai yang rendah. Begitu pula apabila ada satu unit yang memiliki nilai tinggi pada sebuah kueri, dua unit lain cenderung tinggi pula nilainya. Kecenderungan serupa juga terlihat pada *precision at 10* sebagaimana terlihat pada Gambar 5.12.



Gambar 5.12. *Precision at 10* pada 3 unit perolehan informasi.

Metode MAP merupakan diskriminator yang baik untuk membandingkan kinerja antar strategi perolehan informasi. Hasil ujicoba yang dilakukan menunjukkan unit undang-undang memperoleh nilai MAP paling tinggi, diikuti oleh unit bab dan unit pasal. Berdasarkan hasil tersebut dapat dikatakan bahwa unit undang-undang merupakan unit perolehan informasi yang lebih baik ketimbang dua unit lainnya. Sementara metode *precision at n* menunjukkan seberapa banyak dokumen relevan yang diperoleh dalam n dokumen teratas. Dua metode *precision at n*, yakni $p@5$ dan $p@10$, memberikan hasil yang sama. Pasal di peringkat pertama, diikuti oleh bab dan undang-undang.

5.5 Analisis Hasil Ujicoba

Bagian ini memaparkan hasil analisis terhadap ujicoba yang telah dilakukan, terutama hasil evaluasi sistem.

5.5.1 Analisis Unit Undang-undang

Unit pengindeksan undang-undang memiliki hasil p@5 0,44, p@10 0,25, dan MAP 0,94. Unit undang-undang selalu memiliki dokumen relevan pada 10 hasil dokumen teratas. Dalam merancang kueri, penulis membaca terlebih dahulu beberapa undang-undang baru kemudian menentukan kueri yang akan dipergunakan. Dengan demikian, muncul dugaan bahwa dalam setiap kueri akan ada unit undang-undang yang relevan karena kueri dibuat dengan mengambil intisari dari undang-undang tersebut.

Berdasarkan Gambar 5.1 ada kasus dimana undang-undang memperoleh nilai MAP yang minim, yakni pada kueri ke-36. Pada kueri ke-36 penulis menginginkan sistem menampilkan dokumen yang membahas tentang hukuman pidana penjara dan denda bagi pelaku pencemaran nama baik. Kueri yang penulis buat yakni “pidana penjara denda pencemaran nama baik”. Hasilnya hanya ada sebuah dokumen yang relevan dengan kebutuhan informasi tersebut. Dokumen relevan tersebut berada di peringkat keempat. Peringkat pertama sampai ketiga didominasi oleh undang-undang yang membahas masalah pencemaran lingkungan hidup dan hukuman pidana penjara bagi pelakunya. Hal tersebut terjadi karena minimnya dokumen yang membahas topik dalam kueri dan dominannya topik yang hampir serupa dengan kueri.

Kueri-36 juga menunjukkan salah satu keunggulan unit undang-undang dibanding unit pasal dan unit bab, yakni keluasan cakupannya. Dalam metode penulisan undang-undang, pasal-pasal yang memuat permasalahan yang sama akan dikelompokkan dalam bab yang sama pula. Kueri ke-36 memuat dua permasalahan yang berbeda, yakni jenis tindak pidana (pencemaran nama baik) dan sanksi hukumnya (penjara denda). Undang-undang mengelompokkan dua topik tersebut ke dalam dua bab yang berbeda. Apabila pencarian dilakukan ke

unit bab atau pasal, maka dua hal tersebut akan sulit disatukan karena berada pada dua dokumen berbeda. Akan tetapi, apabila pencarian dilakukan ke unit undang-undang, maka peluang berhasilnya akan lebih tinggi karena dua permasalahan tersebut terangkum dalam dokumen yang sama.

Hasil ujicoba menunjukkan jumlah undang-undang yang muncul dalam hasil pencarian untuk topik-topik tertentu cenderung sedikit. Berdasarkan hasil perhitungan *precision at n*, jumlah dokumen undang-undang yang relevan berkisar 2 sampai 3 dokumen saja. Fenomena tentang undang-undang ini dapat dikaitkan dengan penelitian yang dilakukan oleh Noortwijk yang menyebutkan bahwa secara statistik, kata-kata dalam dokumen hukum bersifat tidak universal [NOO96]. Banyak kata-kata unik ditemukan dalam sedikit dokumen saja. Hal ini menandakan jumlah dokumen yang relevan dengan topik tertentu memang sedikit.

Namun demikian, ada pula kasus dimana cukup banyak undang-undang yang memuat suatu topik. Contohnya adalah kueri ke-3 yang membahas arbitrase. Arbitrase sering diikutsertakan dalam undang-undang yang membahas hubungan antar manusia dalam dunia usaha seperti perdagangan, penanaman modal, atau hak cipta. Tidak heran apabila banyak dokumen relevan muncul dalam hasil pencarian untuk kueri tersebut.

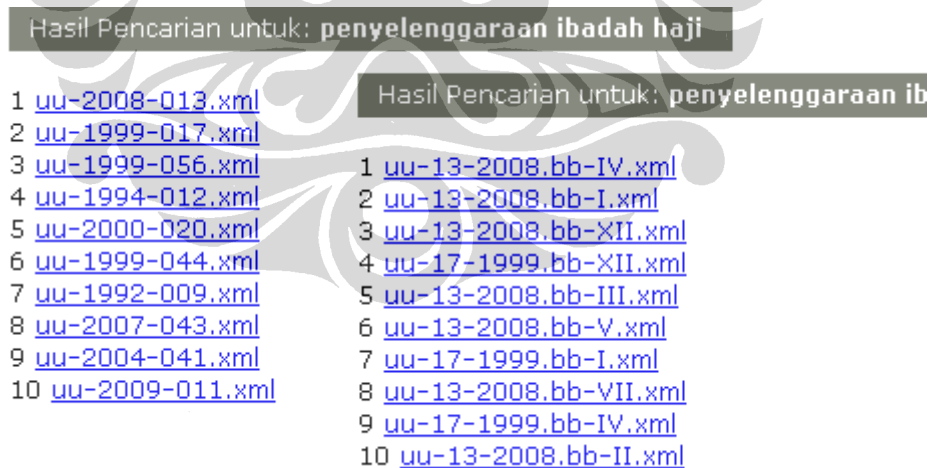
Berdasarkan pembahasan di atas, dapat ditarik benang merah untuk menjelaskan hasil ujicoba unit undang-undang.

- a. MAP yang tinggi menunjukkan kinerja unit ini sangat baik, dokumen relevan sering muncul pada dokumen teratas.
- b. Ada sebagian topik yang dibahas pada banyak undang-undang.
- c. Ada banyak topik yang hanya dibahas pada sedikit undang-undang.
- d. Luasnya cakupan unit undang-undang merupakan faktor yang berharga dalam perolehan informasi.

5.5.2 Analisis Unit Bab

Unit pengindeksan bab memiliki hasil $p@5$ 0.47, $p@10$ 0.33, dan MAP 0.68. Berdasarkan Gambar 5.4, salah satu keunikan unit bab adalah adanya sebagian topik yang tidak memiliki dokumen relevan sama sekali alias MAP-nya bernilai 0. Hal tersebut terjadi karena ada dokumen undang-undang yang tidak memiliki bab dalam strukturnya, sehingga dokumen tersebut tidak terwakili dalam indeks ini. Padahal topik yang dicari bisa jadi ada dalam dokumen tersebut tetapi dirumuskan dalam unit pasal. Ketiadaan elemen bab itu pula yang berpengaruh terhadap minimnya nilai MAP unit bab dibandingkan dengan unit yang lain.

Unit bab memiliki nilai *precision at n* sedikit lebih tinggi dibandingkan dengan unit undang-undang. Artinya ada lebih banyak unit bab yang relevan dengan kueri dibandingkan dengan unit undang-undang. Hal ini terjadi karena jumlah undang-undang yang minim tetapi dalam undang-undang tersebut termuat banyak kata dalam kueri. Dampaknya akan cukup banyak dokumen elemen bab yang relevan. Penulis mengambil contoh kueri nomor 48 pada Gambar 5.12. Hasil 10 dokumen teratas dari kueri tentang “penyelenggaraan ibadah haji” tersebut adalah yakni 2 undang-undang, 10 bab, dan 10 pasal yang relevan.



Gambar 5.13. Perbandingan hasil kueri unit undang-undang dan bab.

Pada Gambar 5.13 terlihat perbandingan hasil pencarian unit undang-undang dan unit pasal dengan kueri “penyelenggaraan ibadah haji”. Untuk unit undang-undang, yang relevan hanya urutan pertama dan kedua saja, yakni uu-2008-013 dan uu-1999-017. Sementara untuk unit bab semuanya relevan. Namun yang

menarik adalah semua dokumen bab yang relevan tersebut merupakan bagian dari unit undang-undang yang relevan. Hal tersebut sesuai dengan karakteristik undang-undang, yakni kata-kata unik tersebar hanya pada sedikit dokumen.

5.5.3 Analisis Unit Pasal

Unit pengindeksan bab memiliki hasil $p@5$ 0.52, $p@10$ 0.43, dan MAP 0.72. Perolehan nilai *precision at n* unit pasal merupakan yang tertinggi dibandingkan dua unit lainnya. Diduga hal ini terjadi karena dalam sebuah undang-undang, ada lebih dari satu pasal yang membahas topik yang sama. Apabila ada lebih dari satu undang-undang dengan topik sama, pasal yang relevan dapat lebih banyak muncul. Namun demikian, unit pasal masih memiliki MAP yang lebih rendah dibandingkan dengan unit undang-undang. Undang-undang 0,94 sementara unit pasal hanya 0,72.

Salah satu penyebab hal tersebut masih terkait dengan karakteristik dokumen, yakni persebaran kata-kata unik dalam sedikit dokumen. Dalam dokumen undang-undang dimana kata-kata unik tersebar hampir merata pada semua pasal, proses perolehan informasi untuk topik tertentu menjadi lebih sulit. Penulis mengambil contoh kueri ke-5, yakni tentang “kecelakaan kereta api”. Hasil pencarian pada unit pasal hampir semuanya berasal dari undang-undang yang sama yakni Undang-undang nomor 23 tahun 2007. Semua pasal tersebut memuat frase “kereta api”. Namun hanya dokumen dengan urutan ke-3, 4, 9 dan 10 yang relevan dengan kebutuhan informasi.

Salah satu kelemahan unit pasal yakni metode penulisan isi pasal, terutama pasal-pasal yang menyangkut konsekuensi hukum akibat suatu tindakan. Perhatikan Gambar 5.14 berikut ini.

```

<PASAL NO="45" ID="uu-11-2008.ps-45">
<AYAT NO="1" ID="uu-11-2008.ps-45.ay-1">
Setiap Orang yang memenuhi unsur sebagaimana dimaksud dalam <REF ID="uu-
11-2008.ps-27.ay-1"/>, <REF ID="uu-11-2008.ps-27.ay-2"/>, <REF ID="uu-11-
2008.ps-27.ay-3"/>, atau <REF ID="uu-11-2008.ps-27.ay-4"/> dipidana dengan
pidana penjara paling lama 6 (enam) tahun dan/atau denda paling banyak
Rp1.000.000.000,00 (satu miliar rupiah)
</AYAT>

```

Gambar 5.14. Penulisan referensi dalam elemen pasal.

Intisari dari pasal tersebut ada dua, yaitu aksi yang menyebabkan pelanggaran hukum dan konsekuensi hukum akibat pelanggaran tersebut. Dalam pasal tersebut, konsekuensi pelanggaran dituliskan dengan jelas, yaitu “..pidana penjara paling lama 6 (enam) tahun dan/atau denda paling banyak Rp1.000.000.000,00 (satu miliar rupiah)...”. Sedangkan aksi pelanggaran hukum hanya diberikan referensi ke pasal yang membahasnya, yaitu “..sebagaimana dimaksud dalam <REF ID="uu-11-2008.ps-27.ay-1"/>...”. Model penulisan pasal yang demikian tentu tidak sesuai dengan kueri yang memuat aksi pelanggaran hukum dan konsekuensinya, seperti kueri-36: “pidana penjara denda pencemaran nama baik”. Padahal pasal-pasal seperti ini tidak sedikit jumlahnya dalam undang-undang Republik Indonesia.