

Studi Mengenai Digital Watermarking pada Dokumen Teks dan Aplikasinya dalam Dokumen Microsoft Word

Gerard Edwin Theodorus (13507079)

Program Studi Teknik Informatika Institut Teknologi Bandung
Alamat : Jl. Ganesha 10, Bandung
e-mail: geedatz@gmail.com

ABSTRAK

Digital watermarking merupakan salah satu cara untuk melindungi hak cipta pada hasil karya seseorang. Digital watermarking dapat dilakukan pada objek tulisan, gambar, suara, maupun video. Yang akan dibahas dalam makalah ini adalah teknik-teknik digital watermarking yang dapat dilakukan pada dokumen teks.

Digital watermarking adalah suatu teknik steganografi yang menambahkan watermark (sesuatu yang dapat menunjukkan identitas pemilik, dapat berupa tulisan atau gambar) ke dalam objek hasil karya pemilik.

Berbeda dengan digital watermarking pada gambar atau suara, penyimpanan data watermark pada bit-bit data dari dokumen akan mengubah isi dokumen, karena itu teknik watermarking pada dokumen teks cukup berbeda. Beberapa contoh teknik watermarking yang digunakan pada dokumen teks adalah Line Shift Coding, Word Shift Coding, atau Character Coding.

Dalam makalah ini, yang akan dibahas adalah perbandingan dari beberapa teknik watermarking yang sudah ada seperti telah disebutkan, cara mendeteksi watermarking pada dokumen teks, jenis-jenis serangan pada watermark, dan contoh aplikasinya secara sederhana dalam sebuah dokumen Microsoft Word dengan menggunakan konsep word shifting.

Kata kunci: *Digital watermarking*, dokumen teks, deteksi, serangan, Microsoft Word

1. PENDAHULUAN

Perlindungan hak cipta merupakan salah satu hal penting. Karena tanpa adanya perlindungan, hasil karya seseorang dapat dengan mudah digandakan dan didistribusikan serta diubah-ubah (*di-edit*). Ada beberapa cara untuk mencegah penyalahgunaan hak cipta, di antaranya adalah :

- Enkripsi
Melindungi data selama proses transmisi. Data hanya bisa diakses oleh pemegang kunci.
- Mekanisme Pencegahan *Copy*
Sulit untuk diterapkan, sehingga jarang digunakan dalam perlindungan hak cipta.
- *Digital Watermarking*
Tidak mencegah pengkopian, tetapi salinan yang ilegal dapat ditelusuri asalnya. Selain itu, dapat menunjukkan data kepemilikan dari suatu data multimedia.

2. DIGITAL WATERMARKING

2.1 Teori Digital Watermarking

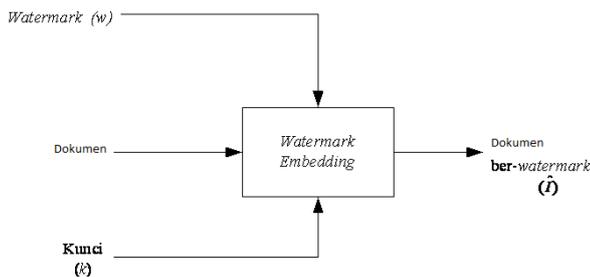
Digital watermarking adalah teknik untuk menandai suatu data multimedia dengan informasi yang dapat menunjukkan identitas pemilik asli dokumen tersebut, misalnya nama atau asal. Fungsi dari digital watermarking antara lain perlindungan hak cipta (*copyright*), pembuktian kepemilikan/*ownership*, serta *fingerprinting* (identifikasi asal data). *Digital watermarking* dapat dilakukan pada data-data :

- Citra → *Image Watermarking*
- Video → *Video Watermarking*
- Audio → *Audio Watermarking*
- Teks → *Text Watermarking*
- Perangkat lunak → *Software Watermarking*

Jenis-jenis watermarking antara lain :

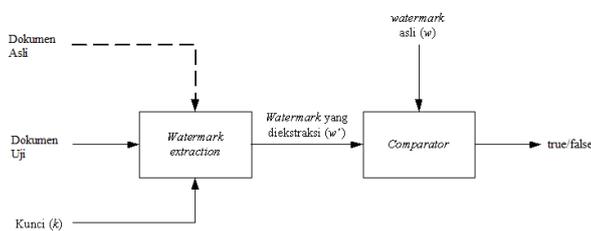
- Fragile watermarking : bertujuan menjaga integritas/orisinalitas media digital
- Robust watermarking : bertujuan menyipkan label kepemilikan media digital

Secara umum prinsip watermarking dapat dilihat dalam gambar berikut :



Gambar 1 Proses embedding watermark untuk dokumen

Dalam proses embedding, data yang akan disisipkan watermark dimasukkan ke dalam embedder dan ditambahkan watermark. Kunci dapat ditambahkan dalam proses ini untuk menambah keamanan.



Gambar 2 Proses extracting watermark untuk dokumen

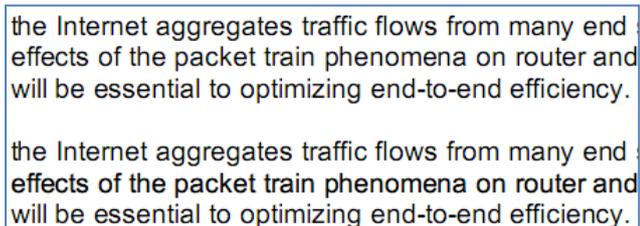
Dalam proses extracting, data yang telah disisipkan watermark dibaca dan dicari watermark yang terkandung dalam data (dapat berdasarkan kunci). Hasil ekstraksi kemudian dibandingkan dengan data watermark dan menghasilkan true/false.

2.2 Text Watermarking

Beberapa teknik watermarking pada dokumen teks yang umum digunakan antara lain :

1. Line Shift Coding

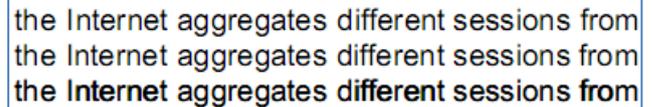
Teknik watermarking dengan mengubah posisi dari suatu baris dalam sebuah teks. Sebuah baris dapat diubah posisinya menjadi lebih atas atau lebih bawah.



Gambar 3 Contoh Line-Shift Coding, baris tengah pada baris paling bawah menggabungkan baris yang di-shift dengan baris aslinya, sehingga terlihat lebih tebal.

2. Word-Shift Coding

Teknik watermarking dengan mengubah posisi kata dalam suatu baris. Sebuah kata dapat digeser posisinya ke kanan dan ke kiri.



Gambar 4 Contoh Word-Shift Coding, baris 1 adalah baris asli, baris 2 beberapa kata di shift ke kiri, dan baris 3 adalah gabungan dari baris 1 dan 2

3. Character Coding

Teknik watermarking dengan mengubah atribut dari sebuah karakter dalam tulisan. Perubahan atribut yang dimungkinkan adalah perubahan ukuran (panjang dan lebar), posisi relatif terhadap karakter lain, maupun perubahan bentuk dari karakter itu sendiri.



Gambar 5 Contoh Character Coding, pada contoh ini, huruf 'r' pada Internet digeser ke bawah posisinya

3. DETEKSI

Deteksi dilakukan terhadap profile dari sebuah dokumen teks. Pertama-tama suatu halaman dokumen diproyeksi menjadi array 2 dimensi dengan indeks (x,y) menunjukkan intensitas pixel pada posisi (x,y) (intensitas = 0/1 pada dokumen hitam putih). Jadi, satu baris pada array menunjukkan 1 baris pixel pada dokumen. Dengan adanya profile, pendeteksian teks akan lebih mudah karena baris array yang tidak mengandung teks akan memiliki intensitas 0 pada semua elemennya. Profile adalah proyeksi dari array 2 dimensi tersebut. Profile tersebut diperoleh dengan cara menjumlahkan elemen-elemen dalam satu baris (horizontal profile) atau dalam satu kolom (vertical profile).

Beberapa teknik deteksi yang dapat dilakukan antara lain feature detection, centroid detection, dan correlation detection.

1. Feature detection

Dalam deteksi ini, posisi dari sebuah kolom diperkirakan berdasarkan posisi dari fitur-fitur yang

ada dalam kolom tersebut. Fitur yang biasanya digunakan adalah sisi yang naik atau turun dari sebuah kolom. Kemudian, fitur tersebut dicari dalam profile yang telah dibuat dan diperiksa lokasinya untuk menentukan apakah fitur tersebut telah bergeser atau tidak.

2. *Correlation detection*

Deteksi dilakukan berdasarkan sinyal yang memiliki tambahan Gaussian noise putih.

3. *Centroid detection*

Ketika efek translasi tidak dapat dilihat secara akurat, *correlation detection* memberikan hasil yang buruk. Deteksi centroid menghitung pergeseran atau delta berdasarkan titik tengah dari suatu profile.

Perbandingan dari teknik-teknik deteksi yang ada :

- Feature detection baik digunakan untuk mendeteksi line shifting karena dapat secara mudah mendeteksi fitur seperti baseline.
- Centroid detection dapat digunakan pada word shifting dan line shifting, tetapi ketika ada noise, centroid detection hanya bagus untuk mendeteksi line shifting.
- Correlation detection memiliki performa lebih baik dalam mendeteksi word shifting, tetapi sangat tergantung pada dokumen asli untuk mengekstrak watermark dalam dokumen.

4. SERANGAN DAN PENCEGAHAN

Watermark yang ditambahkan dalam sebuah dokumen teks dapat dengan mudah dihilangkan dengan cara mengetik ulang keseluruhan daokumen. Karena dengan demikian semua pergeseran baik baris, kata, atau huruf akan hilang dari dokumen. Hal demikian tidak bisa dihindari.

Namun, text watermarking masih dapat digunakan untuk penyebaran dokumen-dokumen lain yang tidak bertujuan komersil, misalnya untuk makalah, artikel koran atau majalah, dan sebagainya.

Jenis-jenis serangan pada text watermark hampir sama dengan jenis serangan pada data multimedia lainnya. Serangan ini dapat berupa perubahan isi dokumen, perubahan pergeseran karakter, baris atau huruf, dan sebagainya.

Hal-hal yang dapat dilakukan untuk mengurangi ancaman akibat serangan dari luar, yaitu :

- Membuat watermark lebih sulit untuk diubah
 - Membuat harga dokumen asli lebih murah dari harga kopi ilegal.
 - Pembuatan watermark yang valid dibuat lebih sulit
- Mode distribusi dari dokumen mungkin dapat membuat penghapusan watermark lebih mudah atau lebih

sulit. Watermark mudah untuk dihapus jika dokumen didistribusikan dalam format bahasa standar seperti PostScript, PDF, dan sejenisnya karena dalam format tersebut informasi baris dan kata dituliskan secara eksplisit dan dapat diubah dengan mudah. Akan lebih sulit untuk mengekstrak watermark dalam bentuk bitmap (dokumen yang ditulis disimpan sebagai gambar bitmap), terlebih ketika teks memiliki banyak jenis dan ukuran font dan berisi banyak tabel dan gambar.

Semakin kompleks atau rumit sebuah dokumen, teknik pendeteksian watermark tidak lagi dapat dilakukan secara otomatis, tetapi membutuhkan intervensi dari manusia. Salinan yang dicetak harus di-*scan* terlebih dahulu menjadi bitmap agar lebih sulit dideteksi, karena bitmap hasil *scan* memiliki lebih banyak noise dibandingkan dengan bitmap asli, dan kemungkinan besar memerlukan intervensi manusia untuk dapat diekstraksi watermarknya.

Satu cara untuk membuat harga salinan ilegal lebih mahal adalah dengan cara membuat file yang disebarakan memiliki bit lebih banyak daripada dokumen asli, misalnya dengan membuat file bitmap dengan resolusi lebih tinggi. Dengan demikian, untuk mempublikasi atau menyalin dokumen yang ilegal akan membutuhkan usaha yang lebih besar.

Selain itu, kemampuan suatu aplikasi untuk mendeteksi salinan yang ilegal juga berguna untuk mencegah serangan terhadap watermark. Misalnya dengan mengecek apakah watermark yang terkandung di dalam dokumen merupakan watermark yang valid.

5. APLIKASI

5.1 Pengantar Microsoft Word

Microsoft Word merupakan salah satu editor dokumen/teks yang sering digunakan. Dalam Microsoft Word sendiri sebenarnya sudah ada beberapa fitur-fitur yang dapat digunakan untuk dapat mengamankan dokumen. Seperti misalnya pemberian signature, enkripsi dokumen, penguncian dokumen agar hanya bisa di-*edit* oleh orang-orang tertentu, atau pemberian *visible watermark* berupa tulisan misalnya SAMPLE, COPY, atau nama dan logo yang menunjukkan identitas pemilik dokumen tersebut.

5.2 Aplikasi Add-In Microsoft Word

Meskipun pada Microsoft Word sudah ada fitur-fitur keamanan seperti yang telah disebutkan pada subbab sebelumnya, penulis mencoba membuat sebuah aplikasi yang merupakan add-in atau aplikasi tambahan pada Microsoft Word. Add-in ini dibuat dengan menggunakan VSTO (Visual Studio For Office) dan dibuat sebagai aplikasi tambahan untuk Microsoft Word versi 2007.

Penulis membuat aplikasi sederhana bernama “Word Shift Watermark” yang merupakan Add-In pada aplikasi Microsoft Word 2007. Aplikasi ini dibuat berdasarkan prinsip Word-Shift Coding yang akan dijelaskan pada subbab berikutnya. Watermark yang digunakan dalam aplikasi ini merupakan watermark yang fragile, watermark akan rusak jika terjadi penambahan atau pengurangan jumlah karakter.

5.2.1 Source Code

Aplikasi “Word Shift Watermark (WSW)” ini dirancang berdasarkan prinsip Word-Shift Coding. Langkah-langkah yang dilakukan dalam algoritma WSW ini adalah :

1. Ubah teks watermark yang dimasukkan ke dalam bit-bit, misalkan teks “Kriptografi” diubah menjadi
“0100101101110010011010010111000001110100011011101100111011100110011000010110011001101001”
2. Sisipkan bit-bit dari hasil perubahan tersebut ke dalam teks dokumen secara merata berdasarkan panjang dokumen.
3. Bit disisipkan dengan cara menambahkan offset pada teks di posisi tertentu. Jika bit yang disisipkan adalah bit ‘0’, maka teks berikutnya dari posisi tersebut digeser / diberikan offset sebanyak 1 fontsize ke kiri, sedangkan jika bit yang disisipkan adalah bit ‘1’, maka teks pada posisi berikutnya akan digeser / diberi offset sebanyak 1 fontsize ke kanan.
4. Simpan panjang dokumen dan panjang teks watermark ke dalam file eksternal untuk keperluan pembacaan watermark.

Potongan source code dari aplikasi ini yaitu :

```
/**
 * Fungsi menyisipkan watermark ke dalam dokumen
 */
private void AddWatermark(object sender, EventArgs e)
{
    // hitung karakter pada dokumen aktif
    ...
    // ubah teks watermark ke dalam bit
    ...

    // set posisi watermark dan sisipkan bit
    for (int i = 0; i < start.Length; i++)
    {
        int pos = (count / start.Length) * (i + 1);
        start[i] = doc.Content.Characters[pos].Start;
        end[i] = doc.Content.Characters[pos].Start;
        Word.Range rng = doc.Range
            (ref start[i], ref end[i]);
        if (bit_wm[i] == '0') str = @"\l 1";
        else if (bit_wm[i] == '1') str = @"\r 1";
        doc.Fields.Add(rng, ref Type, ref str, ref bol);
    }
}
```

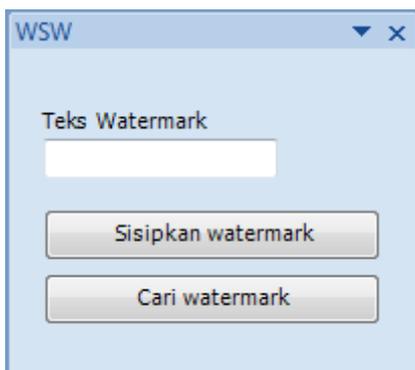
```
// tampilkan pesan berhasil
...
// tulis data-data yang diperlukan dalam file teks
...
}

/**
 * Fungsi mengambil watermark dari dokumen
 */
private void GetWatermark(object sender, EventArgs e)
{
    // baca data-data yang diperlukan dari teks
    ...
    // set posisi pembacaan bit dan baca bit
    for (int i = 0; i < length; i++)
    {
        int pos = (count / length) * (i + 1);
        start[i] = doc.Content.Characters[pos].Start + 2;
        end[i] = doc.Content.Characters[pos].Start + 14;
        Word.Range rng = doc.Range
            (ref start[i], ref end[i]);
        // buat string bit
        if (rng.Text == @"ADVANCE \l 1")
            bit_wm += "0";
        else if (rng.Text == @"ADVANCE \r 1")
            bit_wm += "1";
        else
        {
            MessageBox.Show("Watermark tidak ditemukan");
            bit_wm = "";
            break;
        }
    }
    string watermark = "";

    // konversi bit ke teks
    if (bit_wm != "")
    {
        for (int i = 0; i < length / 8; i++)
        {
            watermark += (char) Convert.ToByte
                (bit_wm.Substring(8 * i, 8), 2);
        }
    }
    // Bandingkan hasil ekstraksi dengan teks watermark
    ...
    // Tampilkan pesan hasil perbandingan
    ...
}
```

5.2.2 Antarmuka

Aplikasi add-in ini muncul sebagai add-in dalam Microsoft Word dalam bentuk task pane, yang muncul di sebelah kanan dokumen dalam Microsoft Word.



Gambar 6 Antarmuka dari aplikasi “Word Shift Watermark”

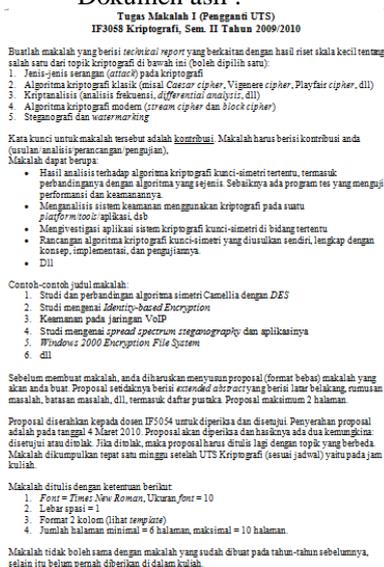
Untuk menggunakan aplikasi ini, pengguna cukup memasukkan teks watermark yang diinginkan ke dalam kotak teks watermark yang telah disediakan. Kemudian, untuk menyisipkan watermark ke dalam teks, pengguna cukup menekan tombol “Sisipkan watermark”. Setelah penyisipan selesai, akan ada pemberitahuan bahwa penyisipan telah berhasil.

Untuk melakukan pencarian watermark dalam dokumen, pengguna cukup memasukkan teks watermark yang sebelumnya dimasukkan kemudian menekan tombol “Cari watermark”. Jika watermark ditemukan dan cocok dengan watermark yang dimasukkan oleh pengguna aplikasi akan memunculkan pesan bahwa watermark telah berhasil ditemukan. Jika watermark tidak ditemukan, aplikasi akan menampilkan pesan bahwa watermark tidak ditemukan.

5.2.3 Hasil Pengujian

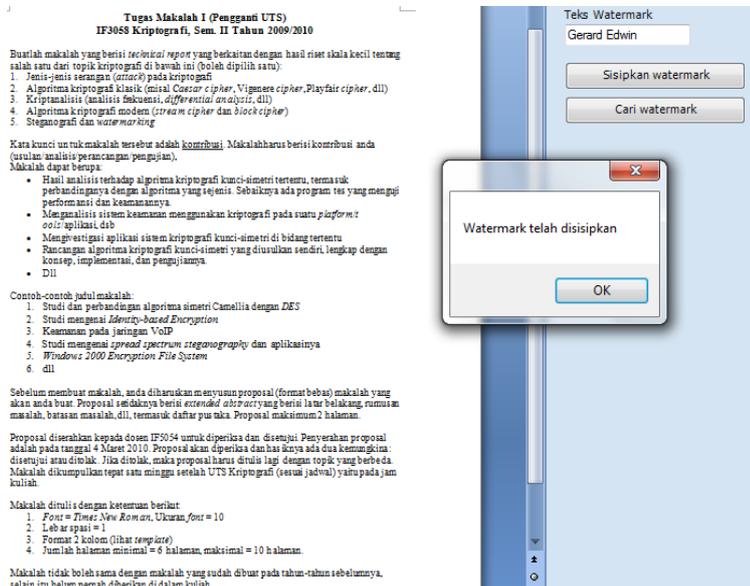
Hasil pengujian dari aplikasi ini dapat dilihat pada contoh berikut :

Dokumen asli :



Gambar 7 Screenshot dari dokumen asli sebelum ditambahkan watermark

Dokumen yang ditambahkan watermark:



Gambar 8 Screenshot dari dokumen setelah diberi watermark “Gerard Edwin”

Jika dilihat secara sepintas, kedua dokumen ini terlihat sama. Tetapi jika dilihat dalam ukuran yang lebih besar, pergeseran karakter dapat terlihat, misalkan pada kalimat berikut :

Kata kunci untuk makalah tersebut adalah kontribusi.

Jika, sebagian dari kalimat tersebut diperbesar, hasilnya adalah :

un tuk
adalah

Gambar 9 Hasil perbesaran sebagian kalimat hasil penambahan watermark

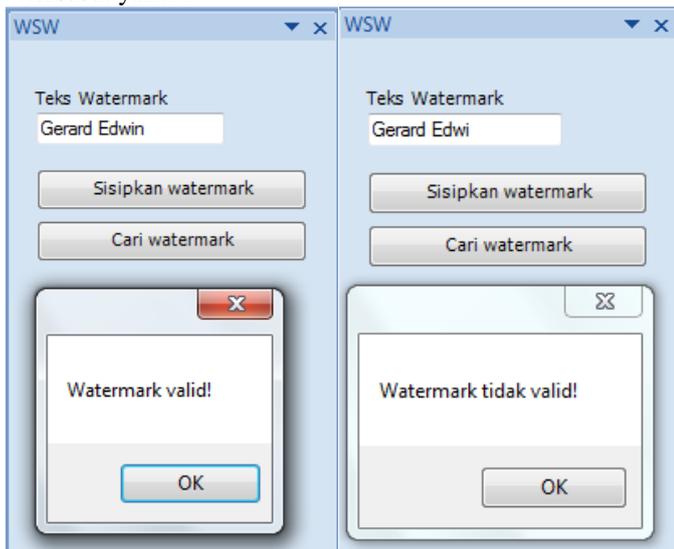
Dapat dilihat pada gambar 7 bahwa jarak antar huruf n dan t dalam kata ‘untuk’ lebih besar 1 fontsize dan jarak antar huruf a dan l dalam kata ‘adalah’ lebih kecil 1 fontsize. Berarti ada bit yang disisipkan di antara huruf-huruf tersebut.

Seperti telah disebutkan, penyisipan bit ini dilakukan secara tersebar di dalam dokumen, sehingga fenomena-

fenomena seperti pada gambar 7 akan tersebar di seluruh dokumen.

Jika orang yang memperoleh kopi dari dokumen ini tidak mengetahui keberadaan watermark, maka orang tersebut tidak akan mencurigai perbedaan spasi antar beberapa karakter, apalagi jika dilihat dalam ukuran yang kecil.

Hasil dari pencarian watermark dalam dokumen tersebut yaitu :



Gambar 10 Kiri : Hasil pencarian dengan watermark 'Gerard Edwin', Kanan: Hasil pencarian dengan watermark 'Gerard Edwi'

5.2.4 Kelebihan dan Kelemahan

Kelebihan dari aplikasi add-in ini adalah :

- Mudah untuk digunakan, karena hanya perlu memasukkan teks yang ingin digunakan sebagai watermark kemudian ditambahkan dengan menekan sebuah tombol saja. Pencarian watermark pun dapat dilakukan dengan menekan tombol "Cari Watermark"
- Dapat digabungkan dengan fitur-fitur keamanan lainnya yang telah disediakan oleh Microsoft Word sehingga dokumen menjadi lebih aman.
- Penambahan atau pengurangan dokumen akan merusak watermark, sehingga aplikasi ini dapat digunakan untuk menguji orisinalitas dari dokumen.

Kelemahan dari aplikasi add-in ini adalah :

- Dalam Microsoft Word, dengan menekan tombol Alt+F9, pengguna dapat melihat semua kode untuk field. Karena word shift dalam aplikasi ini dibuat menggunakan "Advance field", pengguna yang memiliki akses untuk mengubah dokumen dapat

melihat adanya watermark ini dan dapat mengubahnya. Namun, fitur ini tidak diketahui oleh sebagian besar pengguna awam, sehingga aplikasi ini masih cukup aman untuk memberi watermark.

- Tampilan kata yang digeser terlihat ketika format dari paragraph adalah 'justify'. Karena dalam format ini tulisan ditambahkan spasi agar tulisan menjadi rata kiri dan rata kanan. Sehingga kata yang digeser pada bagian tengahnya akan terlihat mencolok.
- Tidak menggunakan key, tetapi pengguna harus memasukkan watermark yang, sehingga dalam hal ini teks watermark dapat berperan sebagai kunci meskipun kurang aman jika dibandingkan dengan penggunaan key.

6. KESIMPULAN

Kesimpulan yang dapat diambil dari makalah ini antara lain :

- Pemberian watermark pada dokumen teks merupakan isu yang penting karena dokumen teks pun seringkali disebarluaskan, sehingga *copyright*-nya perlu dijaga
- Ada beberapa teknik yang dapat digunakan untuk menambahkan watermark pada dokumen teks.
- Selain itu, ada beberapa jenis pendeteksian watermark dalam teks serta beberapa serangan yang dapat dilakukan terhadap watermark dalam dokumen teks.
- Aplikasi add-in yang menggunakan konsep Word-Shift Coding yang dibuat oleh penulis merupakan aplikasi yang cukup aman. Namun, dapat digabungkan dengan fitur-fitur yang disediakan oleh Microsoft Word agar dokumen menjadi lebih aman dari pembajakan hak cipta.

REFERENSI

- [1] Adnan M. Allatar dan Osama M. Alattar, "Watermarking Electronic Text Documents Containing Justified Paragraphs and Irregular Line Spacing"
- [2] Anamitra Makur, "Self-Embedding and Restoration Algorithms for Document Watermark"
- [3] Young-Won Kim, Kyung-Ae Moon dan Il-Seok Oh, "A Text Watermarking Algorithm based on Word Classification and Inter-word Space Statistics"
- [4] J.T Brassil, S.Low, dan N.F. Maxemchuk, "Copyright Protection for the Electronic Distribution of Text Documents"
- [5] S.H. Low, dkk. , "Document Marking and Identification using Both Line and Word Shifting"
- [6] R.G van Schyndel, "A Digital Watermark"
- [7] Bernd Girod, Frank Hartung, Jonathan Su, "Digital Watermarking of Text, Image and Video Documents"
- [8] Rinaldi Munir, "Digital Watermarking"