Abstract: Digital watermarking is the concept of hiding digital signal into digital document such that it cannot be detected and removed easily. That digital signal is called watermark. Text document watermarking is the process of hiding digital data into text document such that it preserves ownership of document. This paper focuses on watermarking process and various techniques including their benefits and limitations.

Index Terms: DWTC, Text Document, Watermarking.

I. INTRODUCTION

In today’s world internet become immense area for sharing data, accessing information, transmission of data from one part of the world to another. On the internet different type of data is available like text, image, audio, video and so on. Now a day’s much educational organization, banking sectors, industries uses internet for their working and transactions. So they have to protect the data from outside world for illegal accessing of information. Protect data from illegal authorities there are various method in digital world like cryptography, steganography, watermarking. In steganography main concept is encryption and decryption of data. In case of encryption data is encrypted and outside world or third parties cannot access it. Then on receiver side after decrypting there is no any provision to protect data from replications. So that there is some other technique is needed in order to prove and give the ownership of data, avoid illegal access to that data, protect it from coping and provide certain type of authentications [1]. The technique came in mind after focusing on above problems is that digital watermarking. The basic idea of digital watermarking is to embed digital information in to digital data such that it cannot be easily detected and removed. The digital information as well as digital data may be text, image, audio, video or combination of anything [2].

1. In visible watermarking the information that is to be embedding in digital data is visible after embedding it in to digital data

2. In invisible watermarking the information that is to be embedding in digital data it cannot seen by normal human eye after embedding it in to digital data.

Watermarking Processes: The process of watermarking involves 4 steps [3].

1. Watermark Generation
2. Watermark Insertion
3. Watermark Detection
4. Watermark Extraction

1. Watermark Generation: In this process there are two possibilities one is direct some unique digital signal as watermark is inserted and another is unique watermark is generated using certain technique.

2. Watermark Insertion: As watermark is generated it is inserted in original digital signal by some insertion process at certain position.

3. Watermark Detection: In this process user over the internet, third parties, illegal authorities, attacker etc just identify or detect that digital document contains watermark or not.

4. Watermark Extraction: This process is performed at recipient side it extract watermark from digital document by using reverse process of inserting watermark.
II. RESEARCH ON TEXT DOCUMENT WATERMARKING

1. Line Shifting: In text document even lines are slightly shifted up or down as a watermark so that normal human eyes cannot detect the difference. Odd lines are use at the time of decoding as a reference [4].

2. Word Shifting: Each line made by number of words. So that even group of words is shifted left or right as a watermark so that normal human eyes cannot detect the difference. Odd group of words is used as reference to calculate the distance at the time of decoding [4].

3. Feature coding: Feature of any specific character is changed in your document. Then it is detected by comparing with original document and watermarked document [4]. Example shows in below table:

<table>
<thead>
<tr>
<th>Character in original document</th>
<th>Character in watermarked document</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>R</td>
<td>R</td>
</tr>
<tr>
<td>J</td>
<td>J</td>
</tr>
<tr>
<td>K</td>
<td>K</td>
</tr>
<tr>
<td>M</td>
<td>M</td>
</tr>
</tbody>
</table>

4. Technique based on not important content: In text document some of the expression change will not change the meaning of the document, so this is used as watermark. Example: Document contains “text, image, audio and video.” So in watermarked document it becomes “text, image, audio, and video.” [5].

5. Synonyms or acronyms based: In original text document specific words selected then it is replaced by its synonyms or acronyms such that meaning of that document should be same. Then resultant document is watermarked one. Example: in original text document contains “This is big city” replace in watermarked document by “This is mega city” or “This is not a small city” [6].

6. Font size and font coding: In font size coding character font size is increased by 0.5 in watermarked document otherwise keep document as it is. In font coding, Microsoft word has many types of fonts some of them are same means its effect is not identified easily. Such fonts are used as watermarking. Example: “Arial” and “Tunga” [7].

7. Key based: It uses object based environment in which each text string is referred as different object. Each object has its own properties and attributes. So watermark is generated using key and embedded based on properties of object like z-ordering of text object [8].
8. Syntactic approach: Normally text document composed of characters, words, sentences. A sentence has different syntactic structure. In order to insert watermark, apply syntactic transformations on text document [1].

9. Semantic approach: Text constitutes verbs, nouns, objectives, prepositions, word spelling, synonyms, acronyms, sentence structure, and grammar rules and so on. Such semantics are used to insert watermark [1].

10. DWTC: This technique based on web documents. Watermark embedded into web document and by using this technique invisibleness and robustness improved [9].

III. PROPOSED WORK

Above mention techniques some are non robust but faster and easier while some are robust but slow and complex, no any technique is ideal one for watermarking. Following are the technique can be proposed:

1. The proposed work can be one image based technique which shifts the paragraph slightly so the security increases as compare to word or line shifting.

2. The proposed work can be technique in which it can perform dual watermarking on document based on the page number means in that document two watermarks are inserted in first two even pages only.

3. As above two techniques involves one or two watermarks still it is not more secure. In order to increase the security inserted multiple watermarks i.e. more than two based on the size of the document in which total size of document is divided in to some equal parts then in each part one watermark is inserted by applying certain conditions.

4. The proposed work can be technique which involves cryptography of watermark as well as invisibleness of watermark.

All mention techniques increase the watermarked document’s robustness, security, watermark capacity, and tamperproof performance.

IV. CONCLUSION

This paper includes the concept of digital watermarking and text watermarking. It gives introductory information of different techniques for text document watermarking. Review on various techniques and its benefits and limitations. Above mention techniques are performs single watermarking expect DWTC is for dual watermarking which give high invisibleness and tamperproof performance. The coding technique in DWTC is can be improved more to increase its performance. This paper also includes the various proposed techniques which can improve the watermarking techniques for more security, robustness, tamperproof performance and so on.

V. REFERENCES


