

# Evaluative Approach towards Text Steganographic Techniques

Sumathy Kingslin\* and N. Kavitha

PG and Research Department of Computer Science, Quaid-E-Millath, College for women, Chennai – 600002, Tamil Nadu, India; sumathykingslin@gmail.com, gnkavita@gmail.com

## Abstract

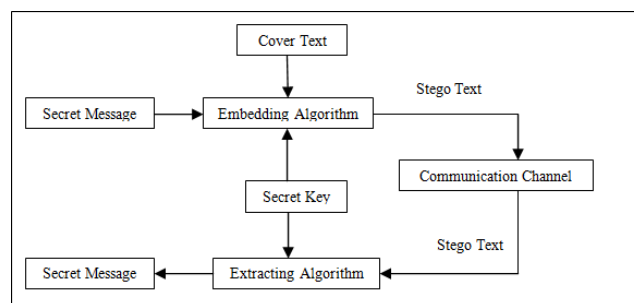
**Objectives:** Steganography is a technique for hiding data in sensitive applications such as banking and legal fields. With Steganography the secret information can be passed without any suspicion. **Methods Analysis:** Text Steganography embed data in a text file through various techniques. Format based Method modifies the cover text to hide the secret message by Resizing the text, Changing the style of text. Text rotation techniques in MS excel document are hard to detect the minor changes in the text's angle through. Mixed case font technique hides one character within each 8 letters of text file. Font Type method use three capital letters to hide one character of secret message. **Findings:** Criteria have been used to measure the goodness of the algorithm. Text rotation method in an excel document benefits the feature of hard to detect and will be applied to excel document only. Its percentage capacity is average and its similarity measure is one. In the mixed case font method payload capacity is very high when compared to other method but it looks like the fonts used in chat rooms. Its percentage capacity is high and its similarity measure is less than one. In Font Type method the embedding capacity is low when compared to other two techniques because only capital letters are used to hide but its similarity measure is one. **Improvements:** Font Type method though has the similarity measure one but it is found that it cannot be applied for numbers and other symbols to hide.

**Keywords:** Capacity Ratio, Font Type Method, Four Criteria, Mixed Case Font Technique, Similarity Measure, Text Rotation Techniques, Text Steganography

## 1. Introduction

Good data hiding hides the data in an undetectable manner in other electronic media or "Covers" like Text, Image, Audio, Video, etc. When data is screened or hidden inside a cover, it is called Steganography and one that employs a text as cover is called Text Steganography<sup>1</sup>. Text Steganography is preferred over other media, because of lesser space occupied by the text, communicate more information and need less cost for printing as well as some other advantages<sup>2</sup>. Each steganography communication system consists of an Embedding algorithm and an Extraction algorithm. The secret message embedded in cover text using Embedding algorithm<sup>3</sup>. Hiding information may require a steganographic key which is additional secret information, such as a password, required for embedding the information<sup>4</sup>. The Embedding algorithm then produces a stego text that can be stored

and/or transferred through communication channels. The Extracting algorithm receives the stego text and the (optional) stego-key, and extracts the secret Message as shown in Figure 1<sup>5</sup>.



**Figure 1.** The main steps of data Hiding and Extracting.

Text Steganography embeds the secret data in text files through various techniques given below:

\* Author for correspondence

### 1.1 Format based Method

It modifies the existing text in order to hide the Steganographic text. It involves the insertion of spaces, resizing the text and changing the style of text to hide the secret message<sup>6</sup>.

### 1.2 Random and Statistical Method

Random Method hides the characters that appear in random sequence. Statistical methods determine the statistics such as means, variance and chi square test which can measure the amount of redundant information to be hidden within the text<sup>6</sup>.

### 1.3 Linguistic Method

Linguistic Method is a combination of Syntax and Semantics methods. Linguistic Steganography considers the linguistic properties of generated and modified text and uses linguistic structure as the space in which messages are hidden. Syntactic steganalysis is to ensure that structures are syntactically correct. Because the text is generated from the grammar, unless the grammar is syntactically flawed, the text is guaranteed to be syntactically correct. In Semantic Method you can assign the value to synonyms and data can be encoded into actual words of text<sup>6</sup>.

## 2. Criteria for Measuring Goodness of Data Hiding Algorithm

---

The technique used for a particular purpose is measured against these basic criteria. No algorithm matches all but a balance should be made before choosing one.

### 2.1 Embedding Capacity

Embedding Capacity (also known as payload) is the amount of data that can be hidden in a cover, compared to the size of the cover. This feature can be measured numerically in units of bit-per-bit (bpb). A Steganographic algorithm with small embedding capacity may have other good features such as Robustness, so it may be the ideal choice when only a small amount of data, such as a short message, has to be hidden<sup>7</sup>.

### 2.2 Invisibility

Any data hidden in a cover causes it to be modified. Invisibility (also termed perceptual transparency or algorithm quality) is a measure of the amount of distortion (alteration) to the cover. A large embedding capacity is useless if it causes large distortions to the cover<sup>7</sup>.

### 2.3 Undetectability

An attacker may be able to detect the presence of hidden data in a given file by computing certain statistical properties of the file and comparing them to what is expected in that type of file<sup>7</sup>.

### 2.4 Robustness

This is a measure of the ability of the algorithm to retain the data embedded in the cover even after the cover has been subjected to various changes as a result of lossy compression and decompression or of certain types of processing such as conversion to analog and back to digital<sup>7</sup>.

## 3. Text Steganographic Techniques

---

Hiding a secret text inside a cover media which is also a text is a tricky one where the cover text file has less redundant bits used for hiding.

### 3.1 Text Rotation Techniques in Ms Excel Document

Convert the secret message to be hidden into Binary bits using ASCII to Binary conversion method. Select the excel document to be used as a cover text. First find the non-empty cell and then find the length of non-empty cell. If the calculated length is less than or equal to limit (limit means number of letters in a cell) specified and if the secret bit is 1 then find that cell contains text or numeric. If it is text then rotate that cell to 1° rotation. Else if it is numeric then rotate that cell to -1° rotation. If the secret bit is 0 leave that cell with no rotation. Finally the formatted excel document is the stego text<sup>8</sup>.

**Embedding Algorithm:**

Input: MS Excel document, Limit p, secret bits

Output: stego-text

Body:

- For each non-empty cell G do.
- Get the selected cell's length n.
- If  $n < p$  and secret bit is 1 then
- If the type of G is text then Rotate the angle of G to  $1^\circ$
- Else If the type of G is numeric then Rotate the angle of G to  $-1^\circ$
- Output the embedded document.

**Extracting Algorithm:**

Input: stego-text, Limit p

Output: secret bits

- For each non-empty cell G do
- Get the selected cell's length n.
- If  $n < p$  then
- If the angle of selected cell is  $1^\circ$  or  $-1^\circ$  angle then Secret bits = 1.
- Else If the angle of selected cell is  $0^\circ$  angle then Secret bits = 0.
- Convert the secret bits into ASCII value and Output the secret message.

**Secret Message:** Arrive On Friday is converted into binary bits as shown in Figure 2.

```
01000001 01010010 01010010 01001001 01010110 01000101 00100000 01001111 01001110 00100000 01000110
01010010 01001001 01000100 01000001 01011001
```

**Figure 2.** Secret Message in ASCII.**Table 1.** Excel Cover Text

No.	Mathematics	English	Physics	Chemistry	Biology	Total	Average
101	90	81	51	8	45	275	55
102	45	44	56	44	45	234	46.8
103	56	44	78	33	56	267	53.4
104	78	33	45	23	78	257	51.4
105	90	81	51	8	45	275	55
106	33	99	55	78	78	343	68.6
107	45	44	56	44	45	234	46.8
108	56	44	78	33	56	267	53.4
109	45	23	56	23	45	192	38.4
110	56	23	77	43	99	298	59.6
111	77	43	33	99	78	330	66
112	33	99	55	78	78	343	68.6
113	56	44	78	33	56	267	53.4
114	45	44	56	44	45	234	46.8
115	56	44	78	33	56	267	53.4

Limit = 4, Cover document = Student marksheet excel document and Font style = MS Serif, size =12

The excel document of student marksheet is used as a cover text to hide secret message as shown in Table 1.

Text rotates in MS Excel document is hard to detect the minor changes in the text's angle through. But when the text's length in a cell is less than 4, the rotation is hard to detect but the stego-text gets worse with the increase of text's length in a cell by setting limit is greater 4.

**3.2 Mixed Case Font Technique in Ms Word Document**

First, the secret message is converted into bits as an array S. The Text file is choosen as a cover text. Each

letter is separately taken in an array T. If ith element of S is bit 1 then the ith element of T is changed to capital letter else if ith element of S is bit 0 then ith element of T is changed to small letter. This method is iterated until the last index element of S is completed<sup>9</sup>.

**Embedding Algorithm:**

Input file: Text file T, Secret Message M.

Output file: Stego Text S.

- Choose a text file T.
- Divide T into letters,  $T=\{T1, T2,----- Tn\}$ .
- Get the secret message M.

**Table 2.** Excel Stego Text

No.	Mathematics	English	Physics	Chemistry	Biology	Total	Average
101	90	81	51	8	45	275	55
102	45	44	56	44	45	234	46.8
103	56	44	78	33	56	267	53.4
104	78	33	45	23	78	257	51.4
105	90	81	51	8	45	275	55
106	33	99	55	78	78	343	68.6
107	45	44	56	44	45	234	46.8
108	56	44	78	33	56	267	53.4
109	45	23	56	23	45	192	38.4
110	56	23	77	43	99	298	59.6
111	77	43	33	99	78	330	66
112	33	99	55	78	78	343	68.6
113	56	44	78	33	56	267	53.4
114	45	44	56	44	45	234	46.8
115	56	44	78	33	56	267	53.4

**A New Digital Image Security Strategy: Steganoflage**  
 Abbas Cheddad, Joan Condell, Kevin Curran and Paul Mc Kevitt  
 School of Computing and Intelligent Systems.

**Figure 3.** Cover Text.

a New digitAl ImaGe sEcUriTy sTraTegY: sTeGaNOfIaGe  
 aBBAs cHeddad, jOan CONDeLL, kEVIn cuRran and Paul MC keViTt  
 sChoOl oF coMpUtInG and IntellIGEnT SysTems.

**Figure 4.** Stego Text.

- Convert secret message M into stream of bits b.
  - Divide b into bits,  $b = \{b_1, b_2, \dots, b_n\}$ .
  - Select  $T_i$  from T and  $b_i$  from b.
  - IF the  $b_i$  is 'one' then change  $T_i$  case into capital else change  $T_i$  case into small.
  - Repeat step 6, 7 till the whole b is hidden.
  - The resultant file will be the stego text S.
- Extracting Algorithm:**
- Get the stego text T, and an empty array S for to store secret bits.
  - Divide T into letters,  $T = \{T_1, T_2, \dots, T_n\}$ .
  - If  $T_i$  case is in capital letter then include bit 1 in  $S_i^{th}$  index else if  $T_i$  case is in small letter then include bit 0 in  $S_i^{th}$  index.
  - Convert the secret bit S into ASCII value to get the secret message.
  - The resultant message will be the secret message.

For the same secret text converted to ASCII secret bits  
 After secret message "Arrive On Friday" is hidden, the

cover text shown in Figure 3 is converted to stego text.  
 Figure 4 shows stego text document.

Hidden information cannot be destroyed when stego text is enlarged or desized. By this method, a large volume of information can be hidden in text when compared with other methods, because the above mentioned method is not using spaces between words or between paragraphs but using the letters themselves. But if they come to know the algorithm or technique the secret information can be easily extracted.

### 3.3 Font Type in MS-Word Document

Before starting this method create a resemble font array which contains a table of cover document font and their resembling fonts for assumption 15 type of cover document fonts and their resembling font. Create a code table that contains coding of each symbol in secret message represented by three types of fonts, thus, 27 characters (English alphabets with space) can be hidden in 3 letters

**Table 3.** Cover Document Font and their Resemblance

Index	Font name		Resemblance	
1	Arial	Arial Unicode MS	Geo_Arial	Arial Narrow
2	Book Antique	Palatino Lino Type	Antique	Caudex
3	Candara	Khmer UI	Ebrima	Microsoft New Tai Lue
4	Century	Century 751 BT	Century Old Style	Century Expd BT
5	Calibri	Gisha	Leelawadee	Liberation Serif
6	Cambria	Proforma	EideticNeoRegular	Liberation Serif
7	Comic Sans	SF Toontime	Komika Text	SF Arch Rival Extended
8	Times New Roman	Tinos	Liberation Serif	Thorndale
9	Helvetica	Arimo	Geo_Arial	Arial-Relcom
10	Courier New	Courier New CE	Tired Of Courier	Tired Of Courier Thin
11	Verdana	Tahoma	Arial	MS sans Serif
12	Perpetua	Chanticleer Roman	Centaur	Caslon Old Face BT
13	Lucida Sans	Lucida Sans Unicode	Segoe UI	Lucida Sans Typewriter
14	Thorndale	Times New Roman	Liberation Serif	Tinos
15	Franklin Gothic Book	Ebrima	Corbel	Trebuchet MS

**Table 4.** Code Table for 27 Alphabets and Space

Index	Characters	F1	F2	F3	Index	Characters	F1	F2	F3
1	A	1	1	1	16	P	2	3	1
2	B	1	1	2	17	Q	2	3	2
3	C	1	1	3	18	R	2	3	3
4	D	1	2	1	19	S	3	1	1
5	E	1	2	2	20	T	3	1	2
6	F	1	2	3	21	U	3	1	3
7	G	1	3	1	22	V	3	2	1
8	H	1	3	2	23	W	3	2	2
9	I	1	3	3	24	X	3	2	3
10	J	2	1	1	25	Y	3	3	1
11	K	2	1	2	26	Z	3	3	2
12	L	2	1	3	27	space	3	3	3
13	M	2	2	1					
14	N	2	2	2					
15	O	2	2	3					

of cover using 3 different fonts, for example: similar font array of Century font is : Century = {Century 751BT, Century Old Style, Century Expd BT}. Secret message is embedded in Capital letters of the cover text document. First step is to find the font of the cover document to get its resemble fonts array. Secondly, scan the capital letters in cover text i.e., needed three capital letters to hide one symbol. Finally, choose the corresponding font type of character in secret message from code table<sup>10</sup>.

- **Embedding Algorithm:**

Open cover document, find its type of font.

Scan cover document to find capitals English letters,

Compute number of capitals English letters to check the capability of embedding.

For each symbol in secret message.

Retrieve its code.

Change font type of three capitals letters by resembling font array according to its code.

- **Extracting Algorithm:**

Each three capitals letters; determine the code of one hiding symbol.

Open Stego document.

For each three capitals letters.

Determine the code.

If the code is (0, 0, 0), then the end of secret.

Message was reached.

Else, find corresponding secret symbol, using code table.

First find the font name of the cover text and then

A New Digital Image Security Strategy: Steganoflage  
 Abbas Cheddad, Joan Condell, Kevin Curran and Paul Mc Kevitt  
 School of Computing and Intelligent Systems, Faculty of Computing and  
 Engineering. University of Ulster, Londenderry, Northern Ireland. United  
 Kingdom  
 Johnny Depp, Robert De Niro, Kevin Spacey, Denzel Hayes Washington,  
 Russell Ira Crowe, John Travolta School of Computing and Intelligent Systems,  
 Faculty of Computing and Engineering. University of Ulster, Londenderry, Northern Ireland. United Kingdom  
 Email: Cheddad\_A@email.ulster.ac.uk

Figure 5. Cover Text.

A New Digital Image Security Strategy: Steganoflage  
 Abbas Cheddad, Joan Condell, Kevin Curran and Paul Mc Kevitt  
 School of Computing and Intelligent Systems, Faculty of Computing and  
 Engineering. University of Ulster, Londenderry, Northern Ireland. United  
 Kingdom  
 Johnny Depp, Robert De Niro, Kevin Spacey, Denzel Hayes Washington,  
 Russell Ira Crowe, John Travolta School of Computing and Intelligent Systems,  
 Faculty of Computing and Engineering. University of Ulster, Londenderry, Northern Ireland. United Kingdom  
 Email: Cheddad\_A@email.ulster.ac.uk

Figure 6. Stego Text.

Table 5. Analyzing the Goodness of the Algorithms

Goodness of the Algorithm	Text rotation techniques in Ms excel document	Mixed case font technique in Ms Word Document	Font Type in MS-Word Document
Embedding Capacity	Short length text in cells in Excel could result in a high embedding rate. As limit specified increase Embedding rate also increases. – <b>VERY HIGH</b> (since many cells in excel document are very short in length, the embedding rate would be very high.)	This approach will insert one character within each 8 letters. So the hiding capacity will be very high compared to other text steganography methods. A large volume of information can be hidden in text compared to other methods - <b>VERY HIGH</b>	We need three characters to hide one character of secret message (one symbol in three capitals letters). This method has good perceptual transparency based on font type’s resemblance. It has <b>VERY HIGH</b> capacity. As secret message increases cover document also rose since capital letters alone are needed to hide.
Invisibility	1° rotation doesn’t make any difference so it is hard to detect through Human vision. <b>HIGH</b>	Alteration depends on number of 1 bit in secret bits stream. - <b>MEDIUM</b>	<b>VERY HIGH</b>
Undetectability	Hidden effect was relatively poor when limit increases. But if Limit is less than equal to 4 it gives high Hidden effect. <b>HIGH</b>	The stego text will attract no attention because it will look like the “cool fonts” used in chat rooms and presentations- <b>HIGH</b>	Hidden effect is <b>HIGH</b> since resemblance fonts are used to hide secret message.
Robustness	<b>HIGH</b>	Hidden information cannot be destroyed when stego text is enlarged or reduced - <b>HIGH</b>	Because the stego document will not change during compression, copying and paste between computer programs, the data hidden in texts remains intact during these operations. <b>HIGH</b>

take the resemblance font name for that font. Cover fonts name and their resemblance name are shown in Table 3.

The coding of each letter in secret message are represented by three types of font, thus, 27 characters (English alphabets with space) can be hidden in 3 letters of cover using 3 different fonts. Table 4 shows the code table.

Secret message = "Arrive On Friday".

Cover document's font style = Verdana, size = 12.

Any word document with more capital letters is chosen as a cover document. Figure 5 shows the cover text.

After secret message "Arrive On Friday" is hidden, the cover text is converted to stego text. Figure 6 shows stego text document.

Six capital letters are needed to hide 2 secret characters. Average ratio of Stego-Documents size increasing in percentage will be 0.766%. But secret message containing numbers and special symbols are cannot be hidden through this method.

## 4. Discussion

### 4.1 Based on the Goodness of the Algorithm

Text rotation in Ms excel document, mixed case font in Ms Word Document, Font Type in MS-Word Document are the techniques differs in the criteria is shown in Table 5.

In Font Type method, to hide two characters of secret message 6 capital letters are needed in cover document. As secret message increases, cover document also increases since capital letters alone are needed to hide. This method is restricted to hide alphabets and spaces only and not for numbers and other symbols. But if the secret message contains only alphabets and spaces this method is the right choice. In the mixed case font method the letter can be hidden in only 8 letters no 8 words as by using spaces between words. Its payload capacity is very high when compared to other method. So, it's a large amount of data compared to other methods keeping the exact meaning of the text and make it looks like the fonts used in chat rooms. Secret message containing Alpha numeric can also be hidden through this method. In text rotation in an excel document benefits the feature of hard to detect. Since many cells in excel document are very short in length, the embedding rate would be very high. Many text-based steganography methods can be used in excel document, text rotation method will be the best method for excel document.

### 4.2 Based on Capacity and Similarity Measure

**Capacity:** Capacity is defined as the ability of a cover text to hide secret message. The capacity ratio is computed by dividing the amount of hidden bytes by the size of the cover text in bytes.

**Capacity ratio = (amount of hidden bytes) / (size of the cover text in bytes).**

Assuming one character occupies one byte in memory, we have calculated the percentage capacity which is capacity ratio multiplied by 100<sup>11</sup>.

Percentage Capacity of Text Rotation techniques in Ms Excel document is 3.38 %, for Mixed case font technique in Ms Word Document is 10.25 %, for Font Type in MS-Word Document technique is 2.90%.

**Jaro-Winkler Distance for Similarity Measure:** The Jaro-Winkler score (or distance) takes into account the number of matching characters and the transposition of characters in two strings. If the Jaro score is 0 then the two strings are dissimilar and 1 means both are exactly same. Jaro score nearest to 1 indicates cover text and Stego text is closely similar. The number of matching (but different sequence order) characters divided by 2 defines the number of transpositions<sup>12</sup>.

In matching the characters, characters can't be more than floor (max (length (s1), length (s2)) / 2) - 1 spaces apart from each other.

To get the Jaro score,

$$1/3 * (m / \text{length}(s1) + m / \text{length}(s2) + (m - t) / m),$$

Where, m is the number of matching characters, s1 is the first string, s2 is the second string, t is the number of transpositions.

To calculate Jaro - Winkler Distance,

$$\text{Jaro\_score} + (L * p * (1 - \text{Jaro\_score})),$$

Where, L is the length of the common prefix at the start of the string up to a maximum of 4, P is the constant scaling factor (usually 0.1 and not more than 0.25)<sup>13</sup>.

In the text rotation technique secret message is hidden without altering the cover text, cover text and stego text are same. So, the Jaro score is 1. In the Mixed case font technique the Jaro score is 0.72 and the Jaro - Winkler distance is 0.77. In the font type technique secret message is hidden without altering the cover text, Cover text and Stego text are same. So, the jaro score is 1. The following Table 6 shows the percentage capacity and Jaro-Winkler distance.

**Table 6.** Comparative Analysis Table

Techniques under discussion	Cover	Works on	Percentage capacity (%)	Jaro Winkler distance
Text rotation techniques in Ms excel document	Student marksheet excel document	Binary value	3.38	1
Mixed case font technique in Ms Word Document	Any English text document	Binary value	10.25	0.77
Font Type in MS-Word Document	Any English text document	ASCII value	2.90	1

## 5. Conclusion

In Table 6 the first technique's embedding capacity is more than third technique and less than second technique. If all the cells in excel document is less than or equal to limit (four) then capacity ratio will be very high compared to the other two techniques. Its Jaro score is 1 so there is no dissimilarity between cover and stego text. In the second technique the embedding capacity is very high when compared to other technique but its jaro score is less than the other two. It shows there is slightly dissimilar with the cover and stego text. In the third technique the embedding capacity is lower when compared to other technique because only capital letters are used to hide. Its Jaro score is 1 which shows the exact similarity between Cover and Stego text.

## 6. References

- Vidya G, Preetha RH, Shilpa GS, Kalpana V. Image Steganography using Ken Ken Puzzle for Secure Data Hiding. *Indian Journal of Science and Technology*. 2014 Sep; 17(9):1403–13.
- Shirali-Shahreza MH, Shirali-Shahreza M. Steganography In Persian and Arabic Unicode Texts using Pseudo-Space and Pseudo Connection Characters. *Journal of Theoretical and Applied Information Technology*. 2008; 4(8):682–7.
- Ramalingam M, Isa NAM. A Steganography Approach over Video Images to Improve Security. *Indian Journal of Science and Technology*. 2015 Jan; 8(1):79–86.
- Kumar A, Pooja KM. Steganography - A Data Hiding Technique. *International Journal of Computer Applications*. 2010 Nov; 9(7):19–23.
- Bhattacharyya S, Banerjee I, Sanyal G. A Novel Approach of Secure Text based Steganography Model using Word Mapping Method (WMM). *International Journal of Computer and Information Engineering*. 2010; 4(2):96–103.
- Mahajan S, Singh A. A Review of Methods and Approach for Secure Stegnography. *International Journal of Advanced Research in Computer Science and Software Engineering*. 2012 Oct; 2(10):67–70.
- Kumar KA, Pabboju S, Desai NMS. Advance Text Steganography Algorithms: An Overview. *International Journal of Research and Applications*. 2014 Transactions; 1(1):31–5.
- Yang B, Sun X, Xiang L, Ruan Z, Wu R. Steganography in Ms Excel Document using Text-rotation Technique. *Information Technology Journal*. 2011; 10(4):889–93. doi: 10.3923/itj.2011.889.893.
- Ali AA, Al-Hussien S. New Text Steganography Technique by using Mixed-Case Font. *OJCSIT*. 2013; 3(2):138–41.
- Bhaya W, Rahma AM, Al-Nasrawi D. Text Steganography based on Font Type in Ms-Word Documents. *Journal of Computer Science*. 2013; 9(7):898–904. doi:10.3844/jcscsp.2013.898.904.
- Agarwal M. Text Steganographic Approaches: A Comparison. *IJNSA*. 2013 Jan; 5(1):91–106.
- Jaro-Winkler distance. 2015 Jan. Available from: [http://en.wikipedia.org/wiki/Jaro%E2%80%93Winkler\\_distance](http://en.wikipedia.org/wiki/Jaro%E2%80%93Winkler_distance)
- Approximate String Matching. 2015 Jan. Available from: [http://biostat.mc.vanderbilt.edu/wiki/Main/ApproximateString\\_Matching](http://biostat.mc.vanderbilt.edu/wiki/Main/ApproximateString_Matching)