

Pdf Redaction

Kevin Willems
Software developer
@ iText Software



Kevin Willems,
Software developer at
iText Software BVBA





Kevin Willems,
Software developer at
iText Software BVBA

Contents

Chapter 1	5	Activating and Setting Up iPhone
	5	What You Need
	5	Activating iPhone
	7	Syncing iPhone with Your Computer
	11	Disconnecting iPhone from Your Computer
Chapter 2	12	Basics
	12	iPhone at a Glance
	14	Home Screen
	17	iPhone Buttons and Touchscreen
	23	Using the Stereo Headset
	24	Connecting to the Internet
	26	Using iPhone on an Airplane
	27	Charging the Battery
	28	Cleaning iPhone
Chapter 3	29	Phone
	29	Calling and Answering
	34	Visual Voicemail
	37	Adding Contact Information to iPhone
	40	Ringtones
	41	Using iPhone with a Bluetooth Headset or Car Kit
	44	Calling to and from Other Countries
	45	Phone Settings
Chapter 4	46	Mail
	46	Setting Up Email Accounts
	48	Sending Email
	49	Checking and Reading Email
	51	Mail Settings

Contents

Chapter 1	5	Activating and Setting Up [REDACTED]
	5	What You Need
	5	Activating [REDACTED]
	7	Syncing [REDACTED] with Your Computer
	11	Disconnecting [REDACTED] from Your Computer
Chapter 2	12	Basics
	12	[REDACTED] at a Glance
	14	Home Screen
	17	[REDACTED] Buttons and Touchscreen
	23	Using the Stereo Headset
	24	Connecting to the Internet
	26	Using [REDACTED] on an Airplane
	27	Charging the Battery
	28	Cleaning [REDACTED]
Chapter 3	29	Phone
	29	Calling and Answering
	34	Visual Voicemail
	37	Adding Contact Information to [REDACTED]
	40	Ringtones
	41	Using [REDACTED] with a Bluetooth Headset or Car Kit
	44	Calling to and from Other Countries
	45	Phone Settings
Chapter 4	46	Mail
	46	Setting Up Email Accounts
	48	Sending Email
	49	Checking and Reading Email
	51	Mail Settings



What are we going to see

What is redaction

Why should you redact

What should you redact

Methods of redaction

When to redact

Getting redaction right



Redaction is removing privileged, sensitive or protected information from a document.

This is typically done before publishing, distributing or archiving documents.



Kevin Willems,
Software developer at
iText Software BVBA

Wikipedia, the free encyclopedia https://en.wikipedia.org/wiki/

From Wikipedia, the free encyclopedia

Anthony John "Tony" ██████████ is a fictional character and the protagonist in the HBO television drama series *The ██████████* portrayed by James Gandolfini. The Italian-American character was conceived by ██████████ creator and show runner David Chase, who was also largely responsible for the character's story arc throughout the show's six seasons. The character is loosely based on real-life New Jersey mobster Vincent "Vinny Ocean" Palermo (born 1944), a former caporegime (capo) and *de facto* boss of the DeCavalcante crime family of New Jersey. Considered to be the model for the DiMeo crime family, several incidents involving the DeCavalcante were incorporated into ██████████ scripts. Bobby Boniello portrayed ██████████ as a child in one episode and Danny Petrillo played the character as a teenager in three episodes.

In the first season, Tony is a capo in the DiMeo crime family. Between the first and second seasons, he is promoted to street boss, a position he retains until the sixth season; his uncle Corrado "Junior" ██████████ is the official boss up until early in the sixth season, but has little or no actual power. Throughout the series, Tony struggles to balance the conflicting requirements of his actual family—wife Carmela, daughter Meadow, son A. J., and mother Livia—with those of the Mafia family he controls. He often displays behavior traits characteristic of a violent sociopath, but also struggles with depression and is prone to panic attacks. He seeks treatment from Dr. Jennifer Melfi in the first episode and remains in therapy on and off up until the penultimate episode of the series.

Gandolfini garnered enormous praise for his portrayal of the character, winning three Emmy Awards for Best Lead Actor in a Drama Series, three Screen Actors Guild Awards for Best Male Actor in a Drama Series and a Golden Globe Award for Best Actor – Television Series Drama as well as two additional SAG Awards for Best Performance by an Ensemble in a Drama Series.

Contents

- 1 Fictional character biography
 - 1.1 Early life
 - 1.2 Narrated killings committed by Tony ██████████
 - 1.3 As a father



James Gandolfini as ██████████

First appearance	"Pilot" (<i>episode 1.01</i>)
Last appearance	"Made in America" (<i>episode 6.21</i>)
Created by	David Chase
Portrayed by	James Gandolfini Bobby Boniello (as a child) Danny Petrillo (as a teenager)
Information	
Aliases	Tony Uncle Johnny Ronald "Big Ron" F. Spears Don Antonio Kevin Finnerty Mr. Petraglia
Gender	Male
Occupation	Crime boss Waste management consultant for Barone Sanitation Co-owner of Satriale's Meat Market and Bada Bing
Title	Soldier (1982 – 1986) Capo (1986 – season 1) De Facto Boss (season 2 – season 6) Boss (season 6)



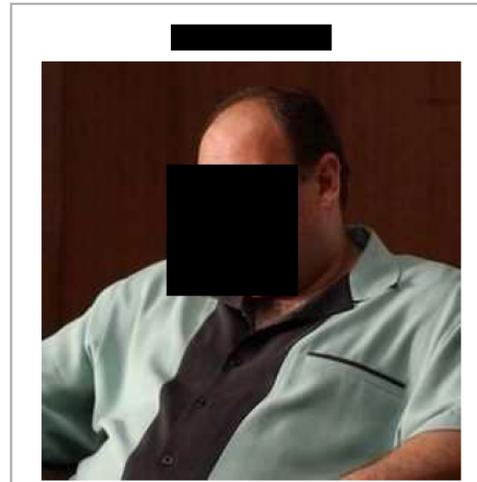
What is redaction?

Ms. Cowley was shot by

on 4:11 pm.

Ms. Cowley was shot by Mr. James on 4:11 pm.

Anthony John "Tony" ██████████ is a fictional character and the protagonist in the HBO television drama series *The ██████████* portrayed by James Gandolfini. The Italian-American character was conceived by ██████████ creator and show runner David Chase, who was also largely responsible for the character's story arc throughout the show's six seasons. The character is loosely based on real-life New Jersey mobster Vincent "Vinny Ocean" Palermo (born 1944), a former caporegime (capo) and *de facto* boss of the DeCavalcante crime family of New Jersey. Considered to be the model for the DiMeo crime family, several incidents involving the DeCavalcantes were incorporated into ██████████ scripts. Bobby Boriello portrayed ██████████ as a child in one episode and Danny Petrillo played the character as a teenager in three episodes.



The screenshot shows a PDF object tree for 'TheSopranos.pdf'. The tree structure includes:

- Info: 1 0 R Modified: Dictionary
- Root: 5 0 R Dictionary of type: /Catalog
 - Dictionary of type: /Catalog
 - Metadata: 19 0 R Stream of type: /Metadata
 - Page: 1
 - PageLayout: /SinglePage
 - PageMode: /UseNone
 - Pages: 2 0 R Dictionary of type: /Pages
 - Dictionary of type: /Pages
 - Count: 1
 - Kids: Array
 - 6 0 R Dictionary of type: /Page
 - Page: 1
 - Contents: 7 0 R Stream
 - MediaBox: Array
 - Parent: 2 0 R Dictionary of type: /Pages
 - Resources: Dictionary
 - ExtGState: 14 0 R Dictionary
 - Font: 16 0 R Dictionary
 - ProcSet: Array
 - XObject: 15 0 R Dictionary
 - Dictionary
 - RG: 11 0 R Stream
 - Stream
 - BitsPerComponent: 8
 - ColorSpace: /DeviceRGB
 - Filter: /DCTDecode
 - Height: 270
 - Length: 6996
 - Subtype: /Image
 - Width: 270

Below the tree is a table of object properties:

| Key | Value | |
|-------------------|------------|---|
| /BitsPerComponent | 8 | ✗ |
| /ColorSpace | /DeviceRGB | ✗ |
| /Filter | /DCTDecode | ✗ |
| /Height | 270 | ✗ |
| /Length | 6996 | ✗ |
| /Subtype | /Image | ✗ |
| /Width | 270 | ✗ |
| / | | ✓ |



Kevin Willems,
Software developer at
iText Software BVBA



Why should you redact?

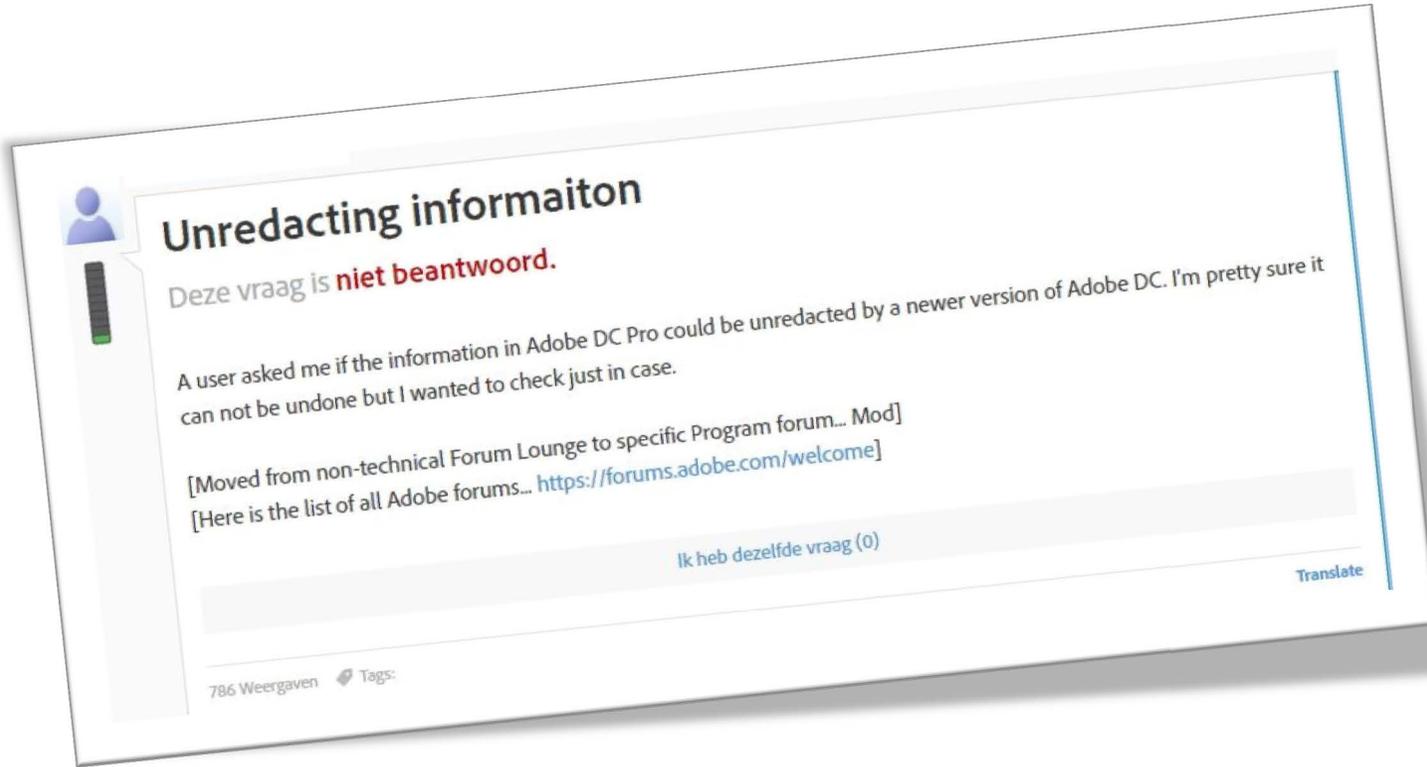
Secrecy protection

- Reduce classification level

Privacy protection

- Guarantee safety and anonymity
- Secure sensitive data
- GDPR Compliance





16-4-2018 Council fined £150,000 for a failure to redact sensitive personal data - Legal Updates - Corporate Commercial Employment Immigra...



The Diverse Law Firm

Legal Updates

Council fined £150,000 for a failure to redact sensitive personal data

16 June 2017 #Employment

🐦 f g+ @ in 📄

Basildon Borough Council have been given a £150,000 fine by the Information Commissioner, after a statement in support of a householder's planning application was published online in full without redacting personal data.

The statement had contained sensitive information about a family's disability requirements including mental health issues, as well as all their names, ages and addresses. It had been posted by an inexperienced council officer who failed to notice the personal information and there was no procedure requiring a second person to check if redaction was required before publishing. The information stayed on line for almost two months before being removed.

The ICO found that, contrary to the Data Protection Act (DPA), the council had failed to take appropriate organisational measures against the unauthorised processing of personal data and made it clear there was no exemption for planning and instead felt the controversial and emotive nature of planning cases made the breach worse.

The size of fine reflected the council's lack of robust procedures to safeguard data and is a warning to all organisations that publish information online that exemptions from the DPA are very limited. Without well thought out procedures and training for staff on data protection, organisations risk facing a similar fine.

Clarkslegal are running a "Getting to Grips with Data Protection" series of webinars, starting on Monday 19 June with a free webinar to address the basic concepts and key principles of data protection. Sign up is free and can be found [here](#).



Art. 17 GDPR
Right to erasure ('right to be forgotten')



Employee and customer data

- Social Security numbers
- Medical information
- Names
- Drivers license or professional license numbers
- Individuals' addresses, dates and months of birth,...
- Other unique identifiers

Company Data

- Financial documents and files
- Proprietary information and trade secrets





Guidance on Redacting Personal Data Identifiers in Electronically Filed Documents

Pursuant to Section ECF-9 of the Administrative Order Regarding Electronic Case Filing, unless the court orders otherwise, parties must refrain from including or must redact the following personal data identifiers from documents filed with the court to the extent required by FRAP 25(a)(5):

- Social Security numbers. If an individual's Social Security number must be included, use the last four digits only.
- Financial account numbers. If financial account numbers are relevant, use the last four digits only.
- Names of minors. If the involvement of an individual known to be a minor must be mentioned, use the minor's initials only.
- Dates of birth. If an individual's date of birth must be included, use the year only.
- Home addresses. In criminal cases, if a home address must be included, use the city and state only.

To ensure compliance with this provision, the court offers the following guidance on improper and proper ways of redacting such information.

COMMON ERRORS WITH REDACTING INFORMATION

A common error in redacting information is to use the wrong method to redact the electronic file. Below is a partial list of misconceptions and methods **NOT** to use:

- Changing the text's font to white will make it look as though the words disappear, but they don't! The metadata that remains can be used by tech savvy individuals to reveal the hidden text you were attempting to redact.
- All word-processing programs (such as Microsoft Word, Corel WordPerfect, WordStar, etc.) retain a lot of hidden code (called "metadata") that can contain revision history and other information. This metadata can reveal anything that was contained in the file at any time, even text that was previously deleted or changed, and even if the file was re-saved. This is a useful tool for tracking revisions, but if this information is not purged from the document, anyone can view this information, even after it has been converted to PDF.
- Adobe Acrobat (the full version) has some graphic and "commenting" tools which can



Recommended Approach: CSR Redaction of Privacy Information

5.1 Removal

- Full patient narratives and corresponding forms (e.g., CIOMS) are **removed**.
- Listings of individual patient data are **removed**.
- Any figures (or tables) that include information pertaining to a single individual are either **removed, or are retained** with the subject identification number **redacted** (e.g. individual subject pharmacokinetic plots), provided this does not jeopardize privacy.

Investigators' CVs / biographies are **removed**.

5.2 Redaction

- **Personal Information:** Names, initials, email addresses, phone and fax numbers, academic/organizational titles, and scanned signatures of all those involved in the study are **redacted** for e.g.
 - o patient / subject
 - o site staff, research institution and staff. Members of Institutional Review Boards / Independent Ethics Committees (IRB/IECs), and members of committees such as Independent Data Monitoring Committees and Statistical Data Analysis Committees (IDMCs/SDACs).
 - o vendor / co-development partner staff
 - o company staff (including monitors listed as emergency contacts in the protocol and names included on citations / references for internal company reports).

Any additional personal information (e.g., facial photographs or comparable images) are also **redacted**.

- **Investigator names:** should be **redacted** unless appropriate contractual agreements are in place to retain them.
- **All non-sponsor company names:** research institution, vendor / co-development partner should be **redacted** unless appropriate contractual agreements are in place to retain them or the partnership is publicly known.
- **Addresses:**
 - o All patient addresses should be **redacted** in full.
 - o All sponsor addresses should be **retained** in full.
 - o All other addresses (investigator, research institution, vendor / co-development partner, IRB/IECs, and IDMCs/SDACs) should be **redacted**, retaining **country**.
- **Patient or subject ID numbers** are **redacted**.
- **Individual outcomes:** ID numbers and the associated description of an individual patient's medical outcome are redacted (e.g., outcome of pregnancy, rare disease or event, congenital abnormality) or sensitive data" (e.g. illicit drug use or "risky behavior") within the body of the CSR, or in a table footer. Further redaction of the outcomes should be assessed by the individual company based on the risk of re-identification.



Kevin Willems,
Software developer at
iText Software BVBA



Page location/region

Regular Expressions

Artificial Intelligence

Manual



Redaction using regex

Lorem ipsum dolor sit amet 078-05-1120,
consectetur adipiscing elit. In arcu turpis,
tempor vitae sodales sed, rutrum cursus
nibh. Sed sagittis tortor sit amet
condimentum 457-55-5462 ullamcorper.
Ut eget diam et nunc porta pharetra.
Etiam cursus vel sem ut maximus. Cras
633-66-2916 vel ligula vel nisi dignissim
gravida sed vel sapien.

Lorem ipsum dolor sit amet [REDACTED],
consectetur adipiscing elit. In arcu turpis,
tempor vitae sodales sed, rutrum cursus
nibh. Sed sagittis tortor sit amet
condimentum [REDACTED] ullamcorper.
Ut eget diam et nunc porta pharetra.
Etiam cursus vel sem ut maximus. Cras
[REDACTED] vel ligula vel nisi dignissim
gravida sed vel sapien.



Kevin Willems,
Software developer at
iText Software BVBA

2018-05-14

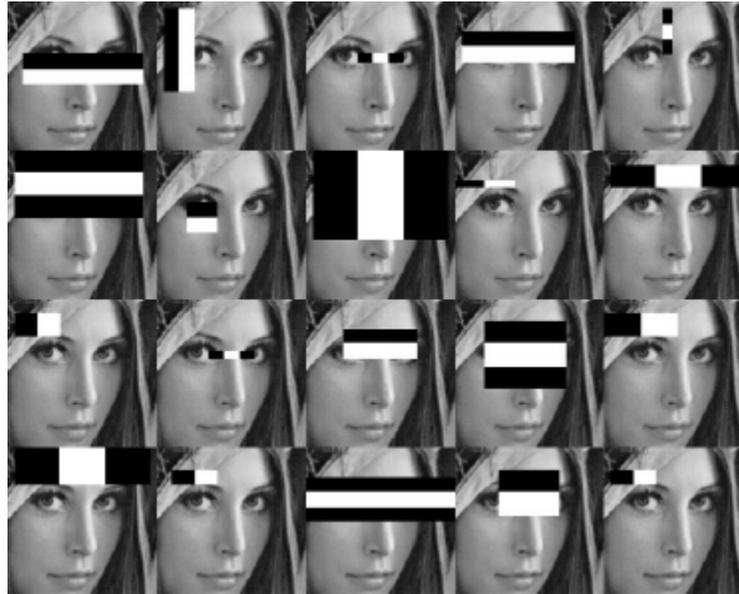
Social Security Numbers: xxx-xxx-xxxx

Credit Card Numbers: xxxx xxxx xxxx xxxx or 3xxx xxxxxx xxxxx

Email addresses: xxxx@xxxx.com/org/net/(other TLDs)



Face recognition



Semantic analysis

Tony Soprano

From Wikipedia, the free encyclopedia

Anthony John "Tony" Soprano is a fictional character and the protagonist in the HBO television drama series *The Sopranos*, portrayed by [redacted]. The Italian-American character was conceived by *Sopranos* creator and show runner David Chase, who was also largely responsible for the character's story arc throughout the show's six seasons. The character is loosely based on real-life New Jersey mobster Vincent "Vinny Ocean" Palermo (born 1944), a former caporegime (capo) and *de facto* boss of the DeCavalcante crime family of New Jersey. Considered to be the model for the DiMeo crime family, several incidents involving the DeCavalcantes were incorporated into *Sopranos* scripts. Bobby Boriello portrayed Soprano as a child in one episode and Danny Petrillo played the character as a teenager in three episodes.

In the first season, Tony is a capo in the DiMeo crime family. Between the first and second seasons, he is promoted to street boss, a position he retains until the sixth season; his uncle Corrado "Junior" Soprano is the official boss up until early in the sixth season, but has little or no actual power. Throughout the series, Tony struggles to balance the conflicting requirements of his actual family—wife Carmela, daughter Meadow, son A. J., and mother Livia—with those of the Mafia family he controls. He often displays behavior traits characteristic of a violent sociopath, but also struggles with depression and is prone to panic attacks. He seeks treatment from Dr. Jennifer Melfi in the first episode and remains in therapy on and off up until the penultimate episode of the series.

Gandolfini garnered enormous praise for his portrayal of the character, winning three Emmy Awards for Best Lead Actor in a Drama Series, three Screen Actors Guild

| | |
|---|---|
| Tony Soprano | |
|  | |
| [redacted] as Tony Soprano | |
| First appearance | "Pilot" (<i>episode 1.01</i>) |
| Last appearance | "Made in America" (<i>episode 6.21</i>) |
| Created by | David Chase |
| Portrayed by | [redacted]
Bobby Boriello (as a child)
Danny Petrillo (as a teenager) |
| Information | |
| Aliases | Tony Uncle Johnny
Ronald "Big Ron" F. Spears
Don Antonio
Kevin Finnerty
Mr. Petraglia |

```
strategy.add(new PhraseCleanupStrategy(new RegexBasedCleanupStrategy("James")))
```





Upon
acquisition

Prior to
distribution

Once the
'work' is
done

Prior to
archival

Prior to
disposal



Getting it right!

Determine your tools

Decide on what to redact

Decide on the method(s) you will use

Decide *when* the data should be redacted

Implement your redaction process

Redaction Process Checklist

This checklist was created to help you gather the information you need to setup your redaction processes. While you'll need to do this for each document containing sensitive data, you can reuse this list to make sure you don't miss a step.

Initial Discovery:

- Call together a few people from different departments. (People generating reports, business users, and legal counsel.)

What to discuss:

- What types of information should be redacted?
- What types of reports might contain this information?
- What are our current processes?
- Are there applicable legal restrictions to redacting specific information?
- Are there applicable legal restrictions to retaining specific information?

Identifying Redactable Content

Where to Start:

- Start with a list of reports or documents that contain information in need of redaction.
- If you need help scanning your current repository, you might start by having someone technical write a script that will find data matching certain patterns like SSN's or your internal account numbers.
- Plug potential knowledge gaps by conducting discovery interviews with end-users. Ask if they have discovered sensitive information in the execution of their job function.

Start Redacting

- Choose a document to start with.
Document Name: _____
- Identify sensitive information in need of redaction.
- Consider your options for how to redact the information.
Circle those that apply: *Page Region, Pattern Matching, Manual*



Redaction
PDF compliance Risk GDPR
Questions?
workflow
<https://itextpdf.com/> iText archive



Kevin Willems,
Software developer at
iText Software BVBA