# I PDF accessibili

Come mai riusciamo a leggere alcuni PDF con Jaws o Zoomtext ed altri no?

La risposta sta nel modo in cui questi documenti sono stati creati. Possiamo suddividere i PDF in 3 tipologie principali che dipendono dal metodo usato per generarli:

1. PDF generati dalla conversione di documenti elettronici

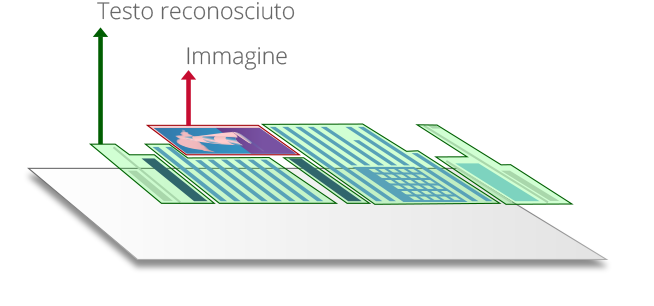
Sono documenti creati da programmi di elaborazione testo o tabelle di calcolo, oppure tramite la funzione di stampa virtuale da qualsiasi altra applicazione. La struttura di questi documenti possiamo paragonarla a quella dei documenti Word, cioè la composizione di testi ed immagini con informazioni sulla loro struttura ed impaginazione. Come nei documenti Word anche qui abbiamo la possibilità selezionare e copiare i testi e quindi leggerli con i nostri programmi di sintesi vocale.

Figura 1: PDF con testi ed immagini su stesso livello. Fonte: www.abbyy.com

1. I PDF scannerizzati o di sola immagine

Questi sono documenti creati ad esempio da stampanti multifunzionali e scanner, o dalla conversione di file immagine in PDF. Anche se si tratta di scansioni di documenti di testo il contenuto di questi PDF è semplicemente un’immagine senza alcun livello di testo sottostante. Se tentiamo di selezionare il testo che vediamo a monitor ci accorgeremo che questo non è possibile. Risulta quindi impossibile anche per le nostre applicazioni d’accessibilità accederne al contenuto.

1. I PDF ricercabili

Quest’ultimi sono principalmente documenti del secondo tipo (quindi solo immagine) a qui viene eseguito l’OCR, cioè il Riconoscimento Ottico dei Caratteri (lo stesso riconoscimento testo che esegue per esempio il programma OpenBook). Una volta riconosciuto il testo presente nell’immagine, questo viene inserito nel PDF in un nuovo livello che va a posizionarsi sotto quello dell’immagine. A differenza quindi del primo tipo di PDF che abbiamo visto, qui non abbiamo testo ed immagini impaginate assieme in un unico livello, ma due starti completamente separati tra di loro. Il primo – l’immagine – usata per la rappresentazione a monitor del documento e il secondo – la trascrizione del testo – non visibile direttamente a monitor, ma che ci permette di selezionarne il contenuto e renderlo anche accessibile alle sintesi vocali.

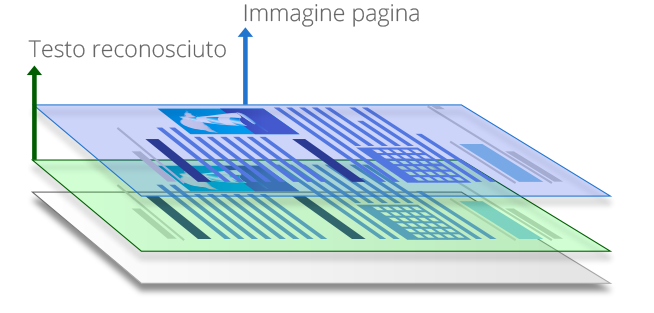


Figura 2: PDF con testi ed immagini su livelli separati. Fonte: www.abbyy.com

In conclusione, sia i PDF generati da documenti elettronici che quelli immagine a cui è stato eseguito il riconoscimento testo saranno leggibili dalle sintesi vocali, ma come potete immaginare con alcune differenze: A differenza dei primi, quelli a cui è stato eseguito l’OCR potranno contenere errori di riconoscimento delle parole. Inoltre l’ordine con cui i vari elementi delle pagine devono venir letti (indicato dalla struttura dei vari elementi tra di loro nei primi) nei secondi sarà rimpiazzato da un più semplice ordinamento dall’alto in basso e da destra verso sinistra (non sempre corrispondente all’ordine di lettura pensato dall’autore del documento, soprattutto nel caso di documenti con complesse strutture tabellari o a colonne).

Fonte immagini e spunti: https://www.abbyy.com/it-it/finereader/types-of-pdfs/