

Introduzione a T_EX4ht ^{*}

Gustavo Cevolani [†]

Versione 0.2
6 febbraio 2005

Sommario

Il presente documento offre una sintetica guida elementare all'uso di T_EX4ht, un convertitore di documenti L^AT_EX in pagine HTML. La discussione è limitata all'uso standard di T_EX4ht e a solo pochi comandi di configurazione avanzata. Anche a questo semplice livello, tuttavia, T_EX4ht è in grado di produrre dalla maggior parte dei sorgenti L^AT_EX documenti HTML più che presentabili per la pubblicazione sul Web.

Indice

1	Introduzione	2
1.1	Come funziona	2
1.2	Installazione	3
2	Conversione automatica: un esempio	3
3	Comandi specifici di T_EX4ht	5
3.1	Opzioni di compilazione	5
3.2	Comandi espliciti	6
	Link e àncore	6
	Codice HTML	7
3.3	Il <i>file</i> di configurazione	7
3.4	Indice analitico	8
4	Un preambolo universale per HTML e PDF	9

^{*}Presentato al *qJrMeeting* 2004 (Pisa, 9 ottobre 2004).

[†]Segnalazioni di errori e commenti in genere sono benvenuti a [gcevolani\[at\]libero.it](mailto:gcevolani[at]libero.it). Desidero ringraziare i fondatori di qJr per l'accoglienza nel gruppo, gli organizzatori del *Meeting* per l'invito e tutti i partecipanti per l'attenzione; Fabiano Busdraghi e Mirko Maischberger per l'aiuto su problemi legati a T_EX4ht e Philip Viton per la fondamentale consulenza sull'installazione (oltre che per il suo [4]).

1 Introduzione

$\text{T}_{\text{E}}\text{X}4\text{ht}$ è un convertitore di documenti sorgenti $\text{\LaTeX}$ in documenti in formato HTML, creato da Eitan M. Gurari e presentato per la prima volta nel 1997.

$\text{T}_{\text{E}}\text{X}4\text{ht}$ non è né un pacchetto $\text{\LaTeX}$ classico (cioè un documento di testo semplice con estensione `.sty`) né un programma separato dal $\text{T}_{\text{E}}\text{X}$: comprende infatti sia un pacchetto vero e proprio (`tex4ht.sty`) che due programmi ausiliari (`tex4ht` e `t4ht`). Richiede quindi un'installazione separata dalla normale distribuzione $\text{\LaTeX}$ (si veda 1.2) ma non può essere usato senza di essa.

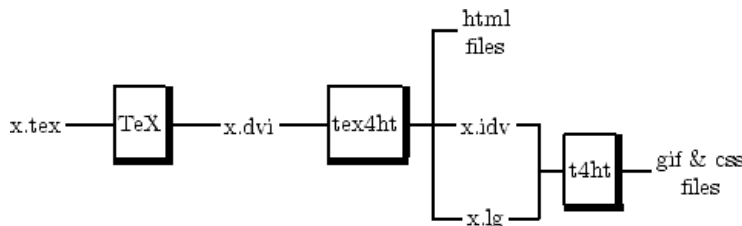
Fra i vantaggi di $\text{T}_{\text{E}}\text{X}4\text{ht}$ si possono annoverare la comodità e la potenza. $\text{T}_{\text{E}}\text{X}4\text{ht}$ fornisce infatti un nuovo compilatore (`htlatex`) che da un sorgente `.tex` produce direttamente come *output* un documento HTML invece che DVI (esattamente come `pdflatex` produce un output PDF). Inoltre non occorre che il sorgente `.tex` venga modificato o opportunamente preparato alla conversione: un qualsiasi documento $\text{\LaTeX}$ viene correttamente tradotto in HTML con risultati più che soddisfacenti dalla compilazione semplice di $\text{T}_{\text{E}}\text{X}4\text{ht}$. In particolare, formule matematiche, immagini e riferimenti incrociati gestiti automaticamente con risultati pressoché perfetti.

1.1 Come funziona

Al contrario di altri convertitori, $\text{T}_{\text{E}}\text{X}4\text{ht}$ opera sul documento DVI prodotto da $\text{\LaTeX}$ per operare la conversione in HTML.

Compilando con `htlatex`, $\text{T}_{\text{E}}\text{X}4\text{ht}$ lancia per tre volte consecutive la compilazione $\text{\LaTeX}$,¹ con un richiamo implicito al pacchetto `tex4ht.sty` (che quindi non è necessario richiamare nel preambolo con `usepackage`; nel paragrafo 4 vedremo che questo è piuttosto comodo). In questo modo si ottiene un output DVI con tutti i riferimenti corretti e contenente una serie di comandi speciali di $\text{T}_{\text{E}}\text{X}4\text{ht}$. Da questo documento `tex4ht`, che è il convertitore vero e proprio, estrae il documento `.html` e due *file* relativi alle immagini (`.idv`) e alle informazioni di stile (`.lg`) necessarie al documento finale. A partire da questi due file, un secondo programma, `t4ht`, crea i file `.gif` necessari e il file di stile `.css`.

L'intero processo è visualizzato nel seguente schema, dove `x.tex` è il file sorgente:



Per una descrizione più approfondita del processo, si veda [1, *The translation process*] e [5, pp. 7–9].

¹Questo significa, fra l'altro, che se la compilazione si blocca su un errore occorre terminarla per tre volte consecutive.

Nota Bene: *in alcuni casi, in verità rari, i file prodotti dalle varie compilazioni interagiscono con nuove compilazioni provocando errori o risultati anomali. Di fronte a problemi, una volta accertato che l'errore non si trovi nel sorgente .tex, eliminare tutti i file di compilazione di T_EX4ht e lanciare la nuova compilazione.*

1.2 Installazione

Nel corso della compilazione, in particolare per produrre le immagini necessarie al documento HTML finale, T_EX4ht deve utilizzare e creare documenti in formato PS e in altri formati grafici, tipicamente GIF.

Per questo, T_EX4ht richiede anche l'installazione di programmi in grado di effettuare le conversioni necessarie. I più comuni sono *GhostScript* (per creare e maneggiare file PS, si veda [8]) e *ImageMagick* (per convertire vari formati immagine fra loro, si veda [9]); entrambi sono liberamente scaricabili dai rispettivi siti.

Philip Viton ha descritto [4] la procedura passo-passo per una installazione completa di un sistema T_EX4ht funzionante in ambiente Windows. Altre informazione generali si possono trovare sul sito di T_EX4ht (si veda [1, *Installation*]).

2 Conversione automatica: un esempio

Una volta installato correttamente (cosa che si può verificare compilando i due *test* inclusi nel programma), T_EX4ht converte automaticamente la maggior parte dei sorgenti L^AT_EX con i risultati voluti.

Si compili ad esempio, lanciando `htlatex prova.tex` da riga di comando, il sorgente seguente, avendo cura di mettere nella cartella di lavoro l'immagine di nome `provafig.jpg` che viene inserita alla riga 31 (o altrimenti commentando questa riga se si vuole rimandare a dopo l'esperimento con la figura):

```
% inizia prova.tex
\documentclass{article}

\usepackage[italian]{babel}
\usepackage{graphicx}

\begin{document}
\title{Prova per \TeX 4ht}
\author{Pinco Pallino}
\maketitle

\tableofcontents

\section{Ciao mondo\dots}
```

```

“Ciao mondo!”, come disse \cite{ciao}.

\section{\dots e ancora ciao.}

Una formula matematica \footnote{Si confronti con%
\emph{The Not So Short Introduction to \LaTeXe}, %
paragrafo 3.1, p. 40.}, nel testo:

$$\lim_{n \rightarrow \infty} \sum_{k=1}^n \frac{1}{k^2} = \frac{\pi^2}{6}$$

e fuori testo:

$$\lim_{n \rightarrow \infty} \sum_{k=1}^n \frac{1}{k^2} = \frac{\pi^2}{6}.$$


Infine un'immagine:

\includegraphics[width=.5\textwidth]{provafig.jpg}

e una breve bibliografia:

\begin{thebibliography}{aaa}
\bibitem{ciao} Pallino, P., \emph{Ciao Mondo},%
Sempronio Edizioni, 2004.
\end{thebibliography}
\end{document}

```

Apprendo ora il documento `prova.html` che è stato creato da `TeX4ht` nella cartella di lavoro, si trova il risultato che ci si potrebbe aspettare da un documento DVI dopo una normale compilazione $\text{\LaTeX}$. Inoltre, le voci dell'indice, le citazioni bibliografiche, i riferimenti (`\ref{...}`) ai paragrafi e i numeri di nota a piè di pagina sono tutti *link* attivi verso i rispettivi destinatari. In particolare, il testo della *footnote* è stato posto in un file HTML separato, chiamato `prova2.html`, a cui il file principale rimanda.²

Assieme al file HTML, nella cartella di lavoro è stata creata una serie di altri documenti, i più importanti dei quali sono:

- `prova.css`, che è il *file di stile* contenente le informazioni di formattazione del documento HTML, secondo le specifiche HTML 4.0;
- `prova.log`, che contiene, oltre alle informazioni di compilazione, tutta una serie di istruzioni supplementari per l'utilizzo di `TeX4ht`;³
- i file `provaN.gif`, con $N \geq 0$, numerati progressivamente e contenenti le immagini (spesso semplici spazi vuoti) necessarie a organizzare graficamente il documento.

²è possibile, come si è fatto nel presente documento, cambiare questo comportamento di *default*: si veda il paragrafo 3.1.

³In particolare se si usa l'opzione di compilazione “`info`”, si veda il paragrafo 3.1.

Naturalmente tutti i file sopra elencati, con l'eccezione di `prova.log`, sono necessari alla corretta visualizzazione del documento HTML; è istruttivo provare a spostare ed aprire il solo `prova.html` in un'altra cartella.

In questo modo, `TeX4ht` gestisce e converte la maggior parte dei documenti `LaTeX`, senza bisogno di intervento da parte dell'utente. Il prossimo paragrafo è invece dedicato alla discussione di alcuni comandi aggiuntivi che permettono di migliorare ulteriormente il risultato.

3 Comandi specifici di `TeX4ht`

Nel seguito, discutiamo brevemente alcuni dei comandi di uso più comune di `TeX4ht`; per una lista più completa e una presentazione più approfondita si vedano [1], [2] e [5].

3.1 Opzioni di compilazione

Esiste una serie di opzioni di compilazione che cambiano il comportamento di `TeX4ht` per vari rispetti. Le opzioni vanno inserite da riga di comando fra virgolette doppie, dopo il nome del sorgente.⁴

Si provi ad esempio a compilare il precedente documento col comando:

```
htlatex prova "html,info,fn-in"
```

Innanzitutto, quando si inseriscono esplicitamente opzioni di compilazione come in questo caso, la prima deve sempre essere relativa a un formato di output (ad esempio `html`, `xhtml`) o a un file di configurazione personale (si veda paragrafo 3.3), altrimenti si ottiene un errore. `html` è l'opzione di default.

Nell'esempio sopra, l'opzione `info` richiede di inserire nel file `log` una serie di informazioni aggiuntive oltre a quelle di compilazione; l'opzione `fn-in` forza `TeX4ht` a mettere le note in fondo al documento HTML e non in pagine separate come prima.

Per documenti di una certa lunghezza, può essere preferibile separare le sezioni una dall'altra: inserendo fra le opzioni i numeri 1, 2, 3 o 4 si produce un documento HTML con una pagina iniziale dedicata all'indice e spezzato in corrispondenza (rispettivamente) del primo, secondo, terzo o quarto livello di intestazione (`chapter`, `section`, ecc.). Si veda ad esempio il risultato della compilazione

```
htlatex prova "html,2,info,fn-in";
```

si noti che le note vengono messe in fondo alla pagina relativa, come è da aspettarsi. I riferimenti ipertestuali fra pagine sono messi automaticamente all'inizio e alla fine di ogni pagina.

Aggiungendo al codice sopra anche l'opzione `frames`, si ottiene un documento HTML diviso in due pannelli verticali, quello a sinistra contenente l'indice fisso

⁴In realtà esistono tre gruppi di opzioni, rispettivamente per `tex4ht.sty` e per i due convertitori. Per altre informazioni si veda [1, *From LaTeX to HTML*].

e quello a destra contenete le altre pagine via via sfogliate, in modo simile alla finestra “Segnalibri” di alcuni documenti PDF.

Un'altra utile opzione è **sections+**, che trasforma ogni titolo di capitolo, paragrafo, ecc. in un *link* verso l'indice iniziale.

3.2 Comandi espliciti

Dei vari comandi specifici di T_EX4ht discuteremo solo quelli che sono forse i più utili immediatamente. Per gli altri si veda [1, 2].

Link e àncore. Il primo serve a creare riferimenti ipertestuali (*link*) a altri documenti, pagine o siti Internet, oltre ai semplici riferimenti interni ottenibili con `\ref{...}`.

Il comando T_EX4ht

```
\Link[file]{pos}{nome} testo \EndLink
```

produce un link di contenuto `testo` a `file#pos`, cioè al punto chiamato `pos` del documento `file`, battezzando `nome` il punto corrente in cui viene richiamato nel documento.

Ad esempio, scrivendo in `prova.tex`

```
\Link[http://www.guit.sssup.it]{}{} GUIT \EndLink
```

si introduce un link “GUIT” al sito in Rete di q_JIr (notare che l'indirizzo deve essere completo dell'intestazione `http://` altrimenti non viene riconosciuto come tale).

Attenzione: negli indirizzi Web, i caratteri `~` (tilde), `%` e `-` devono venire introdotti con i comandi `\string ~`, `\string _` e `\%` rispettivamente.

Se il primo argomento opzionale viene omissso, T_EX4ht si aspetta che sia inserito il secondo: in questo caso il riferimento è al documento corrente, alla posizione `pos`.

Un'àncora (invisibile) all'interno del documento viene dunque prodotta da un comando del tipo

```
\Link[]{}{ciao} \EndLink
```

e viene richiamata da un link del tipo

```
\Link[]{}{ciao}{xxx} ciao \EndLink
```

Notare che tutti gli argomenti di `\Link` sono in effetti opzionali, nel senso che non provocano errori di compilazione se mancano.

Codice html. è possibile, in un documento $\text{\TeX}4\text{ht}$, inserire codice HTML all'interno del comando `HCode`.

Ad esempio, i comandi

```
\HCode{<br>}
\HCode{<hr>}
\HCode{<a name='nome'>} \HCode{</a>}
```

producono rispettivamente un'interruzione di riga, un linea orizzontale e un'ancora da richiamare con “`nome`”. Notare, dall'ultimo esempio, che ogni comando HTML richiede un comando `\HCode` apposito.

Benché comodo, non è conveniente inserire codice HTML direttamente nel documento per non rendere quest'ultimo dipendente dalla compilazione $\text{\TeX}4\text{ht}$ (si veda il paragrafo 4 per questo argomento), a meno che non si stia creando un documento esclusivamente dedicato al Web.

3.3 Il *file* di configurazione

Ulteriori caratteristiche del documento HTML finale possono venir ottenute grazie ai comandi avanzati contenuti in un apposito *file di configurazione* personalizzato. Qui non discuteremo queste configurazioni avanzate se non per un esempio molto semplice; per altre informazioni si vedano [1, *Private Configuration File*] e gli esempi in [5] e [7].

Per i motivi spiegati nel paragrafo 4, conviene mantenere i comandi di configurazione specifici per l'HTML separati dal sorgente del documento.

Per far questo, si crei un nuovo file di testo di nome `prova.cfg` e si scrivano in esso le seguenti righe:

```
% file di configurazione personalizzato
\Preamble{html}
%
\Css{body {font-family: verdana;}}
\Css{body {background-color : yellow;}}
%
\begin{document}
%
\EndPreamble
```

A questo punto è possibile compilare il sorgente $\text{\LaTeX}$ con il comando

```
htlatex prova.tex "prova"
```

cioé richiamando il file di configurazione personale appena creato. In questo modo, il carattere del documento HTML utilizzerà un *font* della famiglia *Verdana* e lo sfondo della pagina sarà colorato in giallo.

Notare che `\Preamble{...}` può contenere vari argomenti separati da virgole, fra cui tutte le opzioni di compilazione considerate nel paragrafo 3.1. Questo permette di compilare il documento senza dover esplicitamente richiamare tutte le diverse opzioni.

3.4 Indice analitico

L'unica caratteristica di base che non viene gestita automaticamente dalla normale compilazione `htlatex` è l'indice analitico.

Benché infatti `TEX4ht` includa correttamente l'indice generato da `MakeIndex`, i numeri di pagina dell'indice non hanno di default riferimenti ipertestuali all'indietro alla citazione relativa nel testo.

Per ottenere anche questo risultato è necessaria una pre-compilazione del file `.idx` fra le due di `htlatex` e `makeindex`.

Si supponga di aggiungere all'esempio `prova.tex` visto sopra le seguenti righe di codice:

```
% inizia prova.tex, con indice analitico
\documentclass{article}
...
\usepackage{makeidx}
\makeindex

\begin{document}
...
\index{GUIT}
...
\printindex
\end{document}
```

in modo da creare un indice analitico.

Per ottenere i numeri dell'indice con link alle citazioni, si lanciano le seguenti compilazioni:

```
htlatex prova
tex '\def\filename{{prova}{idx}{4dx}{ind}} \input idxmake.4ht'
makeindex -o prova.ind prova.4dx
htlatex prova
```

come riportato da `prova.log`.⁵

⁵Su alcuni sistemi i comandi cambiano, richiedendo le virgolette doppie al posto degli apici o anche nessuna virgoletta [*quest'ultima soluzione ad esempio nel caso di chi scrive, con MiK_TE_X 2.4 su WindowsXP Home*]. Si veda anche [7] per una sintassi differente.

4 Un preambolo universale per HTML e PDF

A parere di chi scrive, l'impiego migliore di $\text{\TeX4ht}$, per il quale questo programma riveste una notevole utilità, è la trasformazione di documenti $\text{\LaTeX}$ nati *non* per il Web, ma di cui si renda necessaria o interessante la pubblicazione in Rete. In questo caso, il documento sorgente `.tex` non conterrà comandi espliciti di $\text{\TeX4ht}$ né tanto meno sintassi HTML; sarà quindi sufficiente una compilazione e qualche opzione per ottenere un documento HTML presentabile, come per gli esempi del paragrafo 2 e 3.1.

Tuttavia, anche nel caso si utilizzi $\text{\TeX4ht}$ appositamente per creare pagine Web, e quindi si renda necessario utilizzare i comandi appositi (ad esempio per il riferimenti ipertestuali) discussi nel paragrafo 3.2, riteniamo convenga mantenere al meglio possibile la compatibilità con gli altri compilatori $\text{\LaTeX}$ standard, e soprattutto con `pdflatex`, che offre tutto il supporto necessario agli ipertesti.

Per questo motivo, sull'esempio di [10] da cui è stato preso il codice seguente, conviene utilizzare (come nel caso del presente documento) un preambolo di questo tipo:

```
% preambolo per doppia compilazione HTML/PDF
\ifx\pdfoutput\undefined      % compilazione htlatex
\documentclass{article}
\usepackage{graphicx}
\DeclareGraphicsExtensions{.png, .gif, .jpg}
\newcommand{\href}[2]{\Link[#1]{#2} \EndLink}
\newcommand{\hypertarget}[2]{\Link[]{#1} #2 \EndLink}
\newcommand{\hyperlink}[2]{\Link[]{#1}{#2} \EndLink}
\else
% compilazione pdflatex
\documentclass[a4paper,12pt]{article}
\usepackage[pdftex]{graphicx}
\DeclareGraphicsExtensions{.pdf,.png,.jpg}
\RequirePackage[hyperindex]{hyperref}
\fi
```

In questo modo, lo stesso file sorgente può venire compilato con $\text{\TeX4ht}$ producendo un documento HTML o con `pdflatex` producendo un documento PDF completo di link ipertestuali.

Notare che nel primo caso vengono definiti tre nuovi comandi, `\href`, `\hypertarget` e `\hyperlink` rispettivamente per creare link esterni, àncore interne al documento e link a tali àncore. Essendo omonimi ai comandi del pacchetto `Hyperref`, essi vengono riconosciuti dalla compilazione `pdflatex` e utilizzati correttamente.

L'unico accorgimento necessario è utilizzare cartelle differenti per le due compilazioni, o cancellare i file di compilazione fra una e l'altra: alcuni infatti, come ad esempio i file `.toc`, vengono utilizzati da entrambi i programmi ma in modo diverso, provocando errori.

Riferimenti bibliografici

[→] *Su T_EX4ht:*

- [1] E. M. Gurari, *T_EX4ht: L^AT_EX and T_EX for Hypertext*, pagina principale dell'autore di T_EX4ht, con la descrizione del programma e le istruzioni. <http://www.cse.ohio-state.edu/~gurari/TeX4ht/mn.html>
- [2] E. M. Gurari, *T_EX4ht: Seeded Hooks*, una raccolta di informazioni avanzate di configurazione, ottenibile in formato HTML lanciando `xhlatex mktex4ht` (aprire `mktex4ht.html`).
- [3] E. M. Gurari, *A demonstration of T_EX4ht*, la presentazione di T_EX4ht al TUG97. <http://www.cis.ohio-state.edu/~gurari/tug97/tug97-h.html>
- [4] P. A. Viton, *T_EX4ht under MiK_TE_X*, una guida passo passo all'installazione di T_EX4ht relativa alla più usata distribuzione L^AT_EX per Windows, MiK_TE_X (www.miktex.org). <http://facweb.arch.ohio-state.edu/pviton/support/tex4ht.html>
- [5] F. Popineau, *Affichez vos document L^AT_EX sur le web avec T_EX4ht*, *Cahiers GUTenberg*, 37-38, 2000. Un articolo di presentazione su T_EX4ht, pubblicato sulla rivista del *GUTenberg*, il *Groupe francophone des utilisateurs de T_EX* (<http://www.gutenberg.eu.org/>). <http://www.gutenberg.eu.org/pub/GUTenberg/publicationsPDF/37-popineau.pdf>
- [6] S. Mayer, *L^AT_EX Conversion Example*, una pagina di confronti ed esempi fra differenti convertitori da L^AT_EX a HTML. Si veda anche la pagina relativa a *T_EXConverter*, un'interfaccia grafica gratuita per gestire la conversione di sorgenti L^AT_EX in altri formati (DVI, PDF, HTML) utilizzando vari convertitori standard (<http://www.mayer.dial.pipex.com/tex.htm>). <http://www.mayer.dial.pipex.com/samples/example.htm>
- [7] M. Maischberger, *Da L^AT_EX a HTML con T_EX4ht*, sintetica pagina in italiano di introduzione a T_EX4ht; interessanti anche le altre sezione del sito dell'autore (ad esempio la pagina relativa ai *formati grafici* per la pubblicazione sul Web.) <http://www.lilik.it/~mirko/page.php?pg=tex4ht>

[→] *Altre risorse utili:*

- [8] *GhostScript*, il sito da cui scaricare l'interprete *GhostScript* e il visualizzatore *GhostView/GSView*. <http://www.cs.wisc.edu/~ghost/>
- [9] *ImageMagick*, il sito da cui scaricare il programma di creazione, conversione e visualizzazione di file grafici di vari formati. <http://www.imagemagick.org/>

- [10] M. Pratesi, M. Latini e M. Antonecchia, *L^AT_EX-PDF-Howto*, un articolo molto utile sulla conversione di documenti L^AT_EX in formato PS e PDF. <http://www.telug.it/marco/LaTeX-PDF-HOWTO/LaTeX-PDF-HOWTO.html>
- [11] S. Rahtz, *Hypertext marks in L^AT_EX: the **hyperref** package*, descrizione del pacchetto **hyperref**, necessario per gestire riferimenti ipertestuali in L^AT_EX. <http://www.tug.org/applications/hyperref/manual.html>

Indice analitico

- .cfg, 7
- 1,2,3,4, 5
- HCode, 7
- Link, 6
- fn-in, 5
- frames, 5
- html, 5
- hyperref, 9
- info, 5
- pdflatex, 9
- sections+, 5
- xhtml, 5
- GhostScript, 3
- ImageMagick, 3
- MakeIndex, 8
- \cite, 3
- \footnote, 3
- \index, 8
- \ref, 3
- \tableofcontents, 3

Ancora, 6

Bibliografia, 3

Citazioni bibliografiche, 3

Comandi

- HCode, 7

- Link, 6

Compilazione, 2, 3, 8, 9

- opzioni di, 5

Configurazione

- file di, 7

CSS, 3

Figure, 3

Formule matematiche, 3

HTML, 7

Indice analitico, 8

Installazione, 3

Link, 6

Note a pié di pagina, 3

Opzioni

- 1,2,3,4, 5

- fn-in, 5

- frames, 5

- html, 5

- info, 5

- sections+, 5

- xhtml, 5

PDF, 9

Preamble, 7

Preambolo universale, 9

Riferimenti incrociati, 3, 6

Sommario, 3