

# Greenstone

## Sommario

- ➔ ◆ Esempi di Biblioteche Digitali basate su Greenstone
- ◆ Modalità di ricerca e di browsing in Greenstone
- ◆ Come creare una collezione di documenti
- ◆ Come funziona Greenstone
- ◆ Come personalizzare una collezione di documenti

## Sommario

- ◆ Esempi di Biblioteche Digitali basate su Greenstone
- ◆ Modalità di ricerca e di browsing in Greenstone
- ◆ Come creare una collezione di documenti
- ◆ Come funziona Greenstone
- ◆ Come personalizzare una collezione di documenti

## Caratteristiche generali

- ◆ Greenstone è un sistema di gestione di Biblioteche Digitali sviluppato presso il New Zealand Digital Library Project alla University of Waikato (<http://www.greenstone.org/cgi-bin/library>)
- ◆ Greenstone fornisce servizi per
  - Costruire collezioni di documenti
  - Permettere la ricerca e l'accesso ai documenti

## Tipi di dati gestiti

- ◆ L'informazione viene organizzata in **collezioni composte di documenti**
- ◆ I documenti possono essere rappresentati in una varietà di formati (Word, PDF, postscript, tiff, jpeg, ecc.) e vengono convertiti in una forma standard per la loro indicizzazione.
- ◆ I documenti possono inoltre contenere **testo, immagini, audio, video**, ecc.
- ◆ I documenti non testuali possono essere associati a documenti testuali che ne forniscono la descrizione oppure possono avere una serie di metadati che vengono utilizzati per la ricerca e per il browsing

Pasquale Savino – ISTI-CNR  
Biblioteche Digitali  
Esempi – Il sistema Greenstone

5

## ◆ New York Botanical Garden

The screenshot shows a search interface for the New York Botanical Garden. The search results for 'Bignonia Catalpa' are displayed, including a thumbnail image of the plant. Below the search results, there is a detailed view of the plant's morphology, showing a large leaf with a cluster of small white flowers. The text 'BIGNONIA. Catalpa.' is visible, along with a description of the plant's characteristics and its distribution in the Americas.

Pasquale Savino – ISTI-CNR  
Biblioteche Digitali  
Esempi – Il sistema Greenstone

7

## Biblioteche Digitali che usano Greenstone

- ◆ **Alcuni esempi**
  - Archives of Indian Labour  
→ <http://www.indialabourarchives.org/>
  - Chopin Early Editions  
→ <http://chopin.lib.uchicago.edu/>
  - Lehigh University Digital Bridges Collection  
→ <http://bridges.lib.lehigh.edu/index.html>
  - Mirabilia Vicomercati  
→ <http://www.mirabiliavicomercati.org/sezioni/006/index.html>
  - New York Botanical Garden  
→ <http://www.nybg.org/basc/lib/rolen/page1.html>
  - Project Gutenberg  
→ <http://public.ibiblio.org/gsd/cgi-bin/library.cgi?a=p&p=about&c=qberg>
  - State Library of Tasmania Sheet Music Collection  
→ <http://greenstone.statelibrary.tas.gov.au/>
  - Washington Research Library Consortium Special Collections  
→ <http://www.aladin.wrlc.org/dl/>
- ◆ **The New Zealand Digital Library**  
→ <http://www.saad.uleth.ca/nz/cgi-bin/library>

Pasquale Savino – ISTI-CNR  
Biblioteche Digitali  
Esempi – Il sistema Greenstone

6

## ◆ University of Chicago Library

- ◆ **Chopin Early Editions**

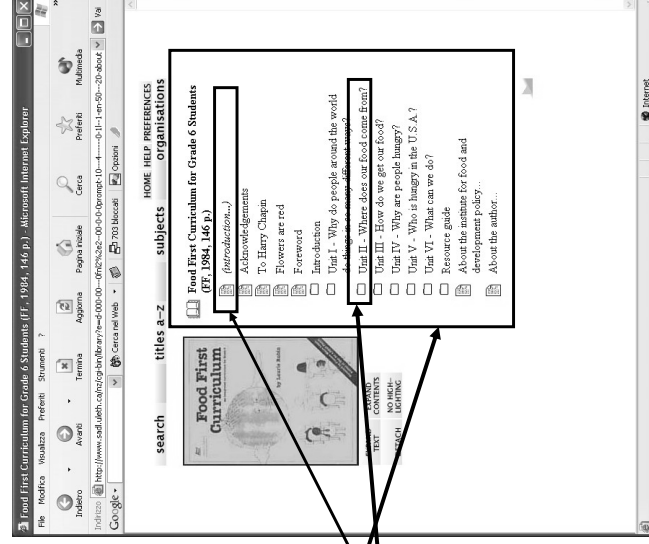
The screenshot shows the University of Chicago Library website for the Chopin Early Editions collection. The top part of the page features a banner with the text 'Chopin Early Editions' and 'The University of Chicago LIBRARY'. Below the banner, there is a search bar and a navigation menu with options like 'search', 'titles', 'uniform titles', 'genres', and 'dedicates'. The main content area displays a list of titles, including 'MÉTHODE DE PIANO FRED. CHOPIN' and 'BALLADE F. CHOPIN'. A detailed view of a musical score is shown, featuring the title 'ADÈLE A. F. CHOPIN. ROMANZESQUE. PREMIÈRE ÉTUDE.' and a page number of 2. The score is displayed in a large, clear font, with a 'view score' button and a 'bibliographic description' button.

P

## Documenti in Greenstone

- ◆ Documenti strutturati (gerarchia)
- ◆ Documenti testuali non strutturati
- ◆ Immagini di pagine di documenti
- ◆ Immagini di pagine e testo riconosciuto via OCR
- ◆ Audio e immagini fotografiche
- ◆ Video
- ◆ Musica

## Modello documenti gerarchico

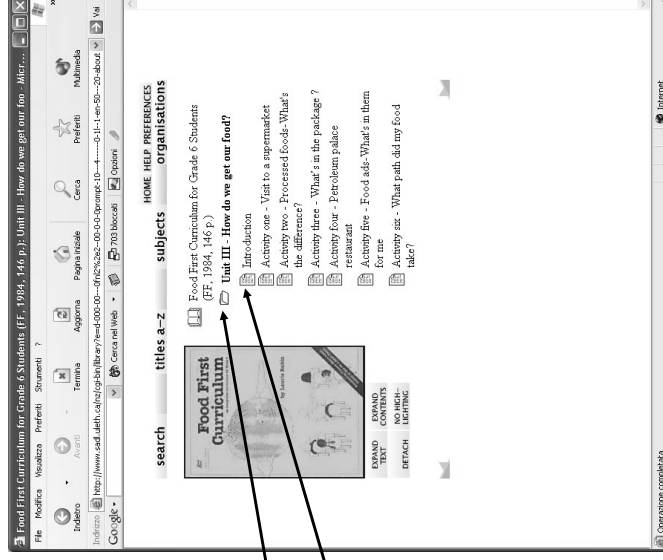


- Documento
  - Sezione
  - Titolo
- I metadati possono essere associati ad ogni livello del documento

## Testo strutturato gerarchicamente

- ◆ Table of contents
- ◆ Capitoli sezioni, sottosezioni, etc.
- ◆ Esempio: [Humanity Development Library](http://www.nzdl.org/)  
→ <http://www.nzdl.org/>

## Modello documenti gerarchico



- Sezione
- Sottosezione

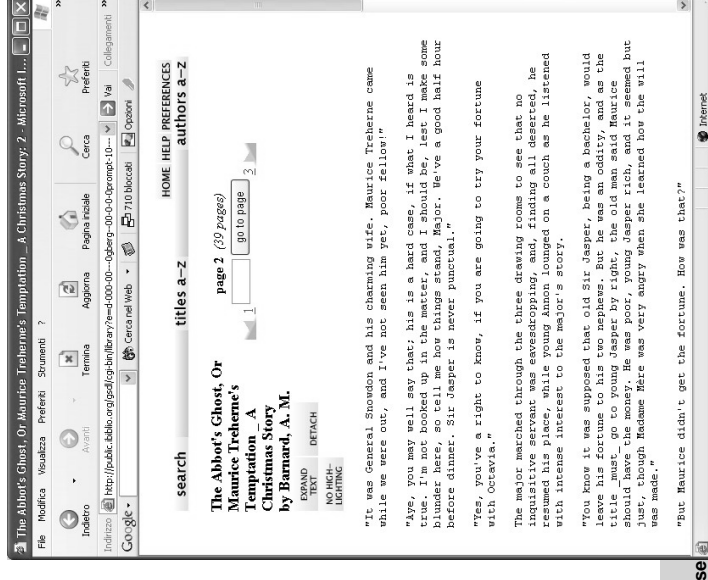
## Testo non strutturato

- ◆ **Testo continuo**
- ◆ **La struttura del documento non è conosciuta dal sistema di biblioteca digitale**
- ◆ **Il Browsing è più difficile in queste condizioni**
- ◆ **Le pagine del documento possono non corrispondere alle pagine fisiche del libro**
- ◆ **Si può avere questa situazione nel caso in cui un libro, un giornale, ecc. sono stati digitalizzati e su di essi è stato fatto OCR**
- ◆ **Esempio: Project Gutenberg Collection**  
→ <http://public.ibiblio.org/gsdll/cgi-bin/library.cgi?a=p&p=about&c=gberg>

## Immagine di pagine di documenti

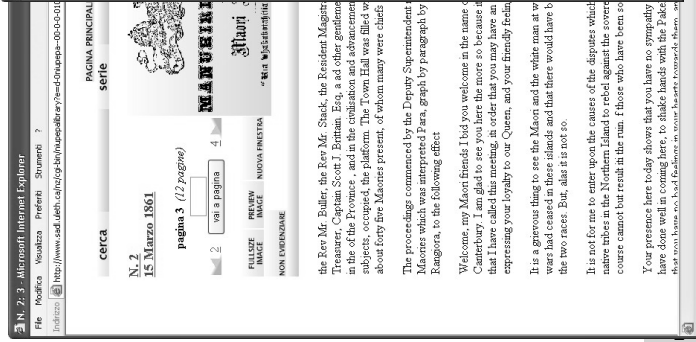
- ◆ **Immagine digitalizzate delle pagine dei documenti**
- ◆ **Accuratezza nel contenuto**
  - OCR è soggetto ad errori
  - È difficile replicare precisamente il formato di presentazione (layout) dell'originale
- ◆ **Maggiori occupazioni di spazio disco**
  - Una immagine digitalizzata richiede circa 20 volte lo spazio disco di una pagina di testo
- ◆ **Richiede maggior tempo per scaricare una pagina dal web**
- ◆ **È necessario fornire una descrizione testuale del documento per la ricerca**
- ◆ **È difficile evidenziare i termini della ricerca sull'immagine**

## Testo non strutturato



## Immagine delle pagine e Testo estratto

- ◆ **Vengono fornite sia le immagini delle pagine che il testo estratto con l'OCR**
- ◆ **La ricerca dei documenti viene effettuata usando il testo estratto**
- ◆ **Vengono visualizzate sia l'immagine della pagina che il testo estratto**
- ◆ **Esempio: Maori Newspaper Collection**  
→ <http://www.saoliuleth.ca/nz/cgi-bin/niupepalibrary?e=d-0niupepa-00-0-0-010-4----full---0-1l-1-it-50---20-about---0001-001-0utzz-8-0&a=p&p=about>



## Altri tipi di documenti

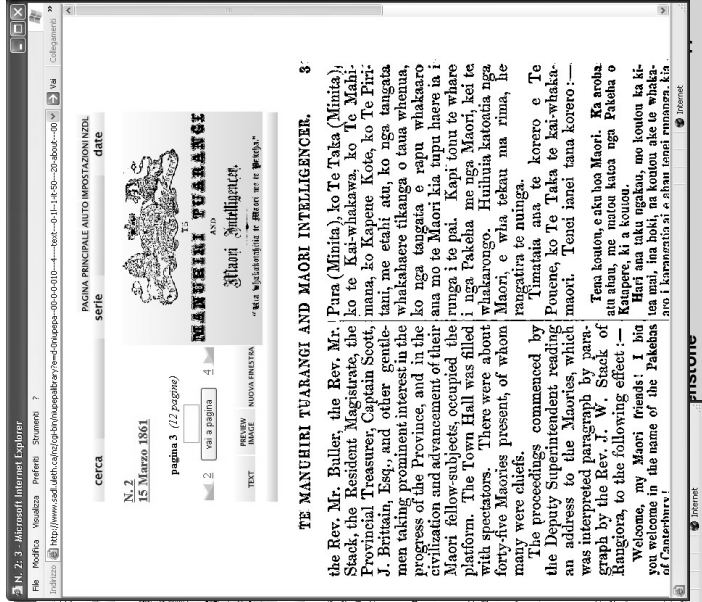
- ◆ **Audio ed immagini fotografiche**
  - Esempio: Oral History Collection, Mirabilia Vicomercati  
→ <http://www.sad.uieuh.ca/nz/cgi-bin/library?a=p&p=about&c=ohist>  
→ <http://www.mirabliavicomercati.org/gsc/cgi-bin/library?a=p&p=about&c=Foto&l=i>
- ◆ **Video**
  - Esempio: [Music Video Collection](http://www.sad.uieuh.ca/nz/cgi-bin/library?a=p&p=about&c=musvid)  
→ <http://www.sad.uieuh.ca/nz/cgi-bin/library?a=p&p=about&c=musvid>
- ◆ **Musica**
  - Varie rappresentazioni: spartiti, MIDI, musica sintetizzata, ecc.  
▪ Esempio: [Music Digital Library](http://www.sad.uieuh.ca/nz/cgi-bin/music/musicalibrary)  
→ <http://www.sad.uieuh.ca/nz/cgi-bin/music/musicalibrary>
- ◆ **Lingue multiple**
  - Interfaccia e documenti
  - Esempio: [Arabic Collection](http://www.sad.uieuh.ca/nz/cgi-bin/library?a=p&p=about&c=arabic&l=ar&nw=windows-1256)  
→ <http://www.sad.uieuh.ca/nz/cgi-bin/library?a=p&p=about&c=arabic&l=ar&nw=windows-1256>

## Biblioteche Digitali

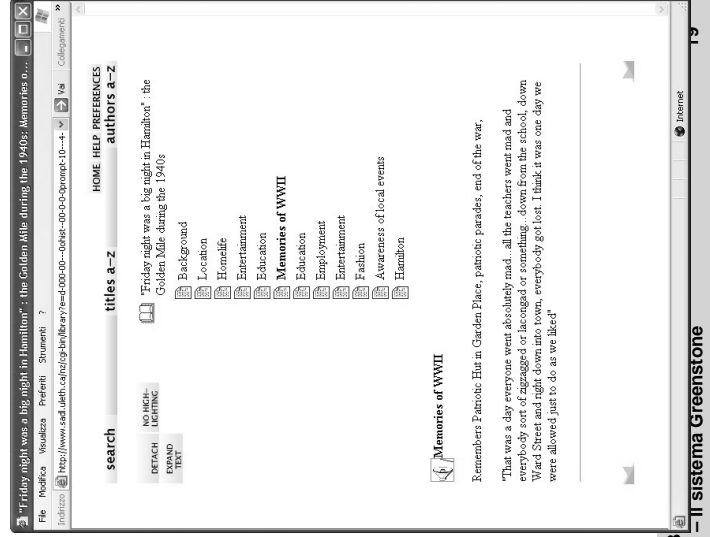
Esempi – Il sistema Greenstone

Pasquale Savino – ISTI-CNR

18



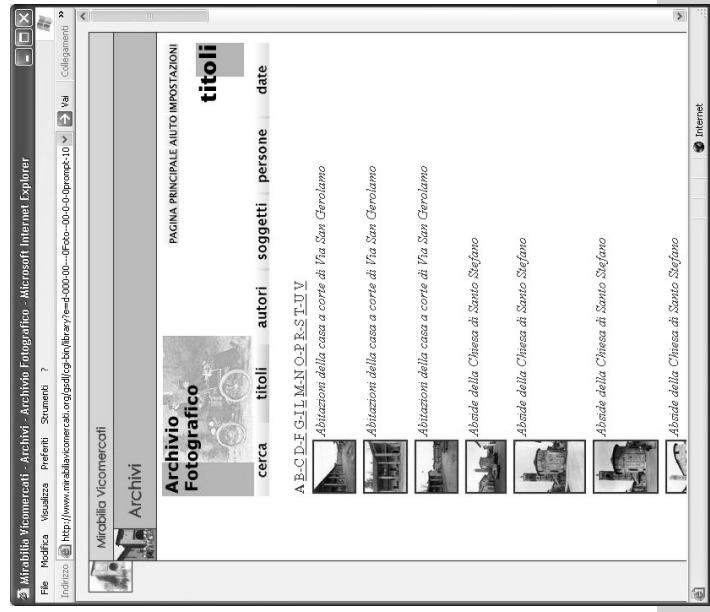
## Audio



Esempi – Il sistema Greenstone

Pasquale Savino – ISTI-CNR

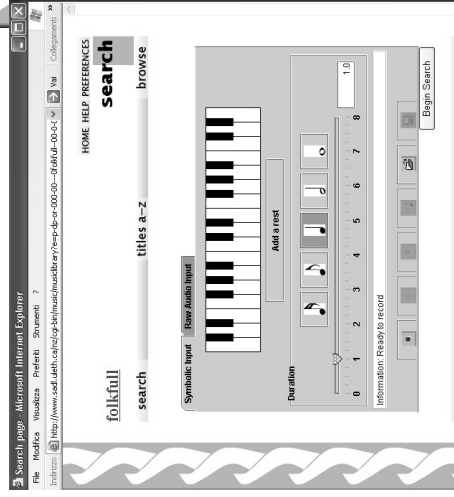
19



## Immagini

Pasquale Savino – ISTI-CNR

## Musica



## Metadati

- ◆ Forniscono informazioni relative ai documenti per facilitarne la ricerca e l'accesso
- ◆ Metadati estratti automaticamente e metadati inseriti manualmente
- ◆ Utilizzo sia di metadati standard che definiti dall'utente
- ◆ I valori dei metadati possono essere strutturati

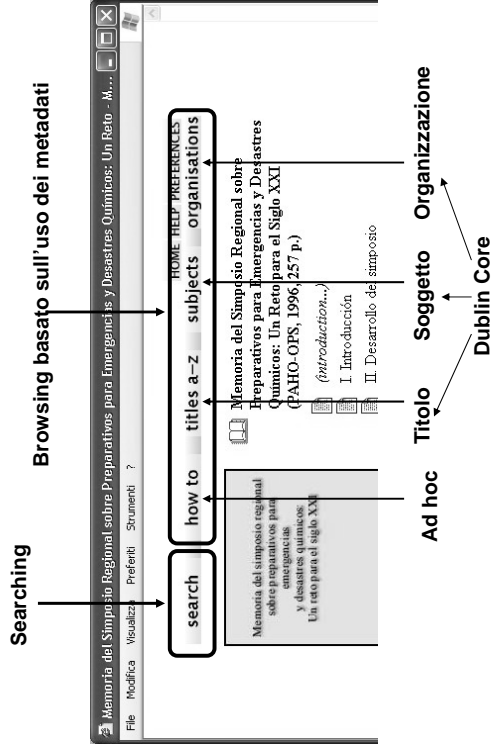
## Esempi di metadati

- ◆ Metadati bibliografici convenzionali
  - Title
  - Author
  - Date
  - Publication
  - Volume Number
  - Issue Number
  - Page Numbers
- ◆ MARC
- ◆ Dublin Core

## Sommario

- ◆ Esempi di Biblioteche Digitali basate su Greenstone
- ➔ ◆ Modalità di ricerca e di browsing in Greenstone
- ◆ Come creare una collezione di documenti
- ◆ Come funziona Greenstone
- ◆ Come personalizzare una collezione di documenti

## Searching and browsing



### ◆ Interrogazioni booleane

- Combina i termini dell'interrogazione con AND, OR, e NOT
- Corrispondenza esatta tra interrogazione e documenti trovati

### ◆ Boolean Queries

- Ad esempio trova tutti i documenti che parlano dell'Italia, che parlano dell'Italia e di ambiente, dell'Italia ma non di ambiente
- Search "Italy", "Italy & Environment", "Italy !Environment"
- Exact match

### ◆ Interrogazioni con ordinamento dei risultati in base alla loro rilevanza (ranking)

- List di termini da trovare
- Corrispondenza non esatta tra interrogazione e documenti ritrovati (ad es. i documenti non contengono tutti i termini specificati)
- Ordinamento dei risultati in base alla loro rilevanza.
- Provatate le stesse query "Italy Environment"

## Searching

- ◆ Tipi di interrogazioni (query) in Greenstone
- ◆ Case-folding e stemming
- ◆ Ricerca di frasi
- ◆ Diverse interfacce di interrogazione

## Case Folding e Stemming

- ◆ Case folding
  - Maiuscole e minuscole trattate allo stesso modo
  - Non è rilevante per alcune lingue
- ◆ Stemming
  - Riduce una parola alla sua radice (ad es. plurale → singolare, riduzione dei verbi all'infinito, ecc.)
- ◆ Non è appropriata per tutte le parti del documento
- ◆ Dipendente dalla lingua

## Ricerca di frasi

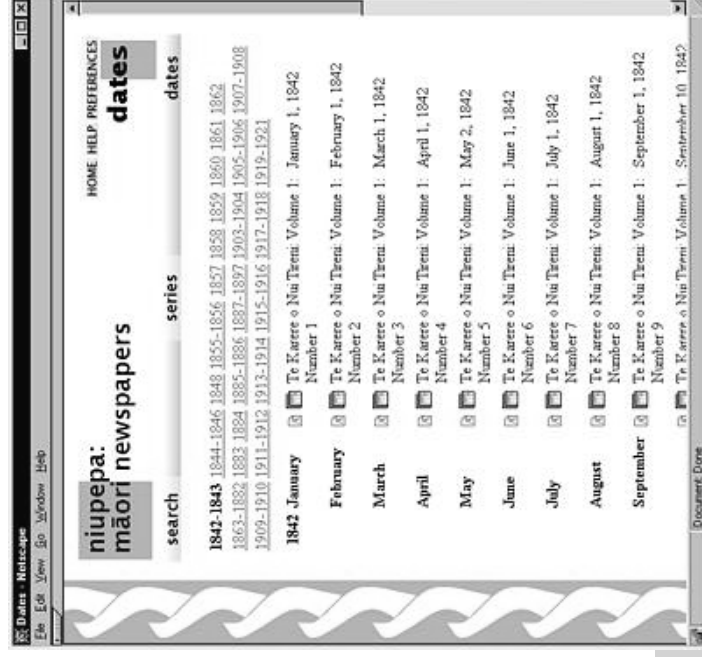
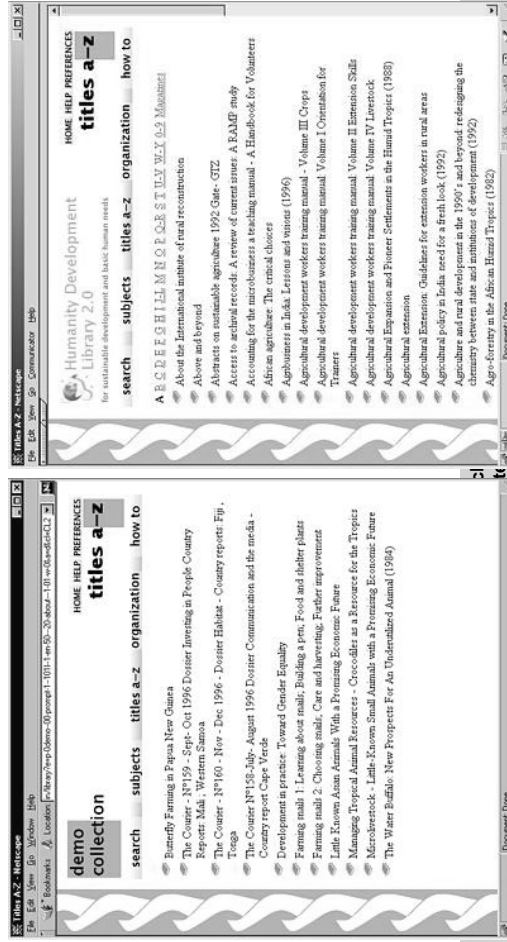
- ◆ Ricerca di un gruppo di parole contigue
- ◆ Due diverse modalità di ricerca di frasi:
  - Postretrieval scan
    - Ricerca i documenti che contengono tutte le parole della frase
    - Determina se i termini sono consecutivi analizzando i documenti ritrovati
    - Richiede lo stesso indice usato per la ricerca delle parole singole, ma le ricerche possono essere lente
  - Word-level index
    - Viene effettuata la ricerca della frase utilizzando un indice specifico che contiene informazioni sulla posizione di ogni parola all'interno di ogni documento
    - Questi indici permettono di determinare se due termini sono consecutivi senza accedere al documento
    - Gli indici sono più grandi di quelli usati per la ricerca di singoli termini, ma la ricerca di frasi risulta più veloce

## Diverse interfacce di interrogazione

- ◆ Ranked o booleana
- ◆ Fielded o non-fielded
- ◆ Case-folding and/or stemming
- ◆ Ranked o con lista dei risultati in ordine naturale
- ◆ Possibilità di usare una search history
- ◆ Attivazione delle diverse opzioni tramite l'uso delle preferenze

## Browsing

- ◆ Basato sui metadati associati ai documenti
- ◆ Browsing di liste alfabetiche (ad es. Titoli, soggetti, ecc.)



## Browsing

- ◆ Browsing per data



## Browsing

- ◆ Browsing basato su strutture gerarchiche di classificazione
- ◆ Ad es. in figura il viene mostrato il primo livello di subject

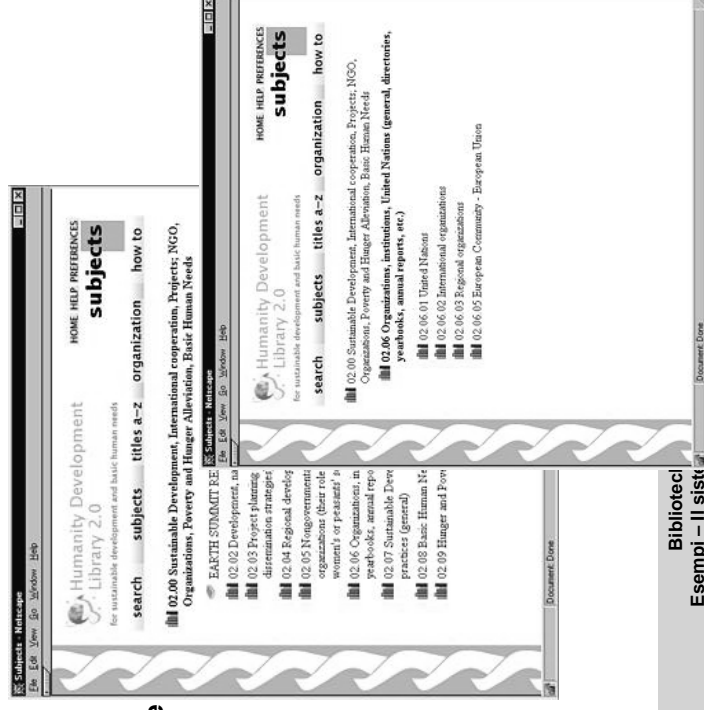


Pasquale Savino – ISTI-CNR

Esempi – Il sistema Greenstone

## Browsing

- ◆ Nelle due figure accanto vengono mostrati i due livelli successivi per subject



Bibliotechi

Esempi – Il sistema

Document Done

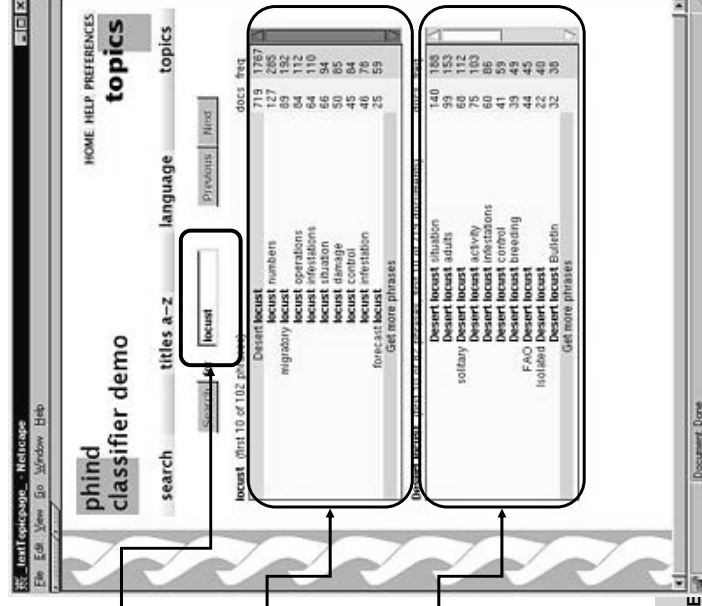
## Browsing su frasi

- ◆ Associare metadati strutturati per la classificazione può essere molto oneroso. Se questa informazione non è disponibile, in Greenstone può essere utilizzato il browsing su frasi presenti nel documento
- ◆ Frase: una sequenza di parole che appaiono più di una volta nella collezione
- ◆ L'estrazione avviene automaticamente
- ◆ Key phrases
- ◆ Browser di frasi
  - Le frasi sono organizzate gerarchicamente
  - Ordinate per documento e per frequenza all'interno della collezione
  - Le foglie di questa gerarchia sono i documenti
- ◆ Esempi: **FAO Collection, The Complete Works of Shakespeare**
  - <http://www.sadi.uleth.ca/nz/cgi-bin/library?a=p&p=about&c=f1998>
  - <http://www.sadi.uleth.ca/nz/cgi-bin/library?a=p&p=about&c=allshake>

Pasquale Savino – ISTI-CNR

Biblioteche Digitali  
Esempi – Il sistema Greenstone

35



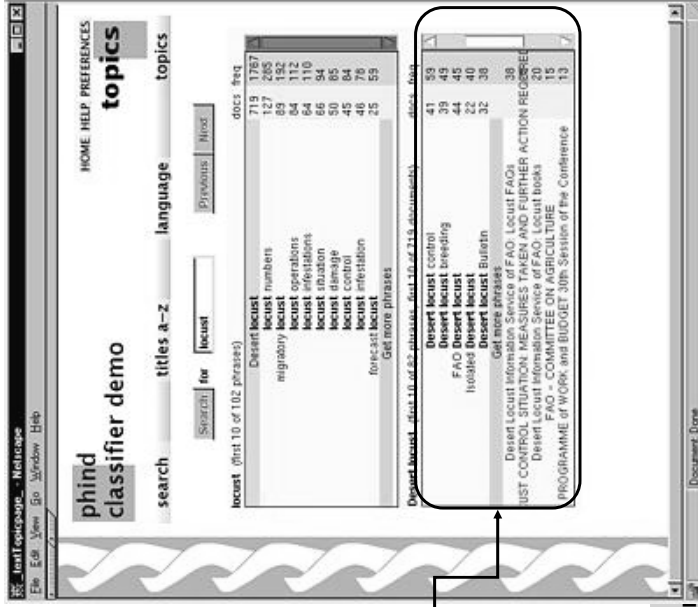
Search frasi che contengono la parola "locust"

Vengono visualizzate le frasi con informazioni sul numero di documenti

Seleziono una delle frasi ("desert locust") e trovo tutte le frasi che la contengono

Document Done

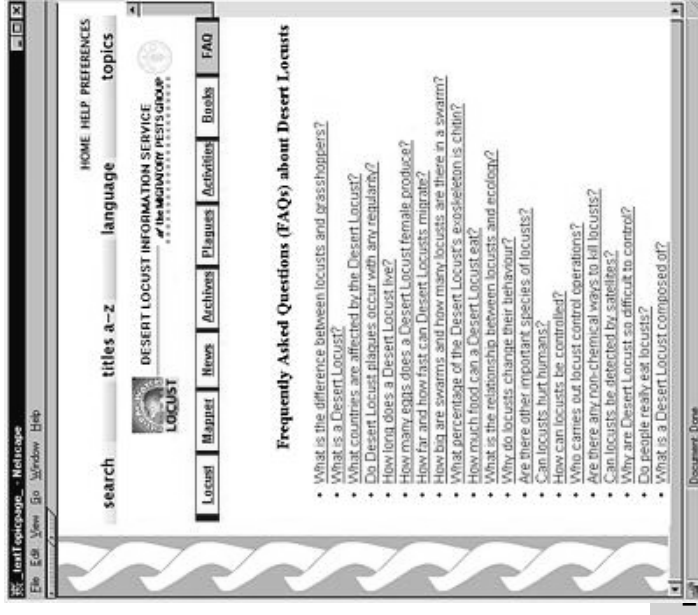
Pasquale Savino – ISTI-CNR



Scendendo nella gerarchia di frasi posso arrivare a documenti che contengono la frase

## Riferimenti

- ◆ Ian Witten, David Bainbridge, “How to Build a Digital Library”, Morgan Kaufmann Publishers, Capitolo 3



Quindi posso visualizzare il documento

## Sommario

- ◆ Esempi di Biblioteche Digitali basate su Greenstone
- ◆ Modalità di ricerca e di browsing in Greenstone
- ➔ Come creare una collezione di documenti
- ◆ Come funziona Greenstone
- ◆ Come personalizzare una collezione di documenti

## Strumenti per creare una collezione

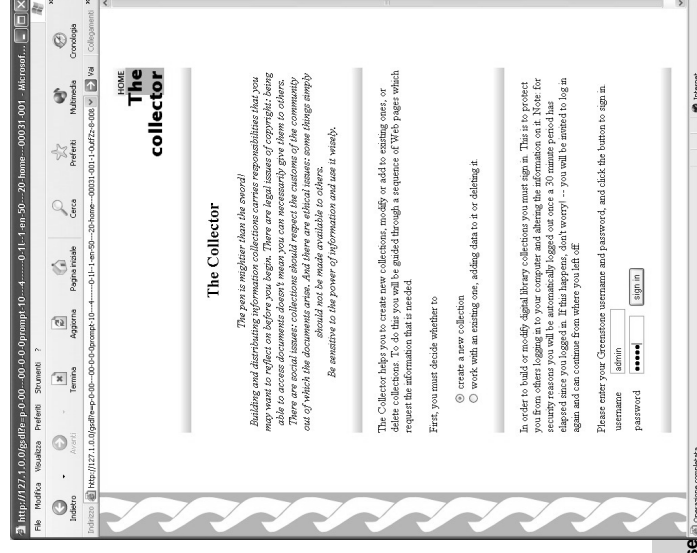
- ◆ **Collector**
  - Interfaccia Web per la creazione delle collezioni di documenti
  - Offre la maggior parte delle funzionalità disponibili
  - Poco adatto se si vogliono eseguire solo alcune funzioni (ad es. rigenerare gli indici)
- ◆ **Creazione manuale**
  - Richiede una conoscenza approfondita del funzionamento del sistema
  - Funziona solo in locale
- ◆ **Uso della Librarian Interface**
  - Funziona in locale
  - Interfaccia per creare nuovi valori di metadati

## Il “Collector” di Greenstone

- ◆ **Il Collector permette di**
  - Creare una nuova collezione di documenti
  - Modificare o aggiungere documenti ad una collezione esistente
  - Cancellare una collezione

## Come avviare il Collector

- ◆ **Click sul link del Collector dalla home page di Greenstone creata dopo l’installazione**
- ◆ **Log in**
  - Quando Greenstone è stato installato, è stato creato un login **admin** con una password scelta durante l’installazione
- ◆ **Il Collector utilizza una interfaccia web standard**



## La home page del collector

## Creazione di una nuova collezione

- ◆ Il compito principale del Collector è quello di creare una nuova collezione
- ◆ La struttura della collezione viene determinata quando la collezione viene definita per la prima volta
- ◆ Il modo più semplice di operare consiste nel copiare la struttura di una collezione esistente e di modificarla a seconda delle proprie esigenze

## Passi per la costruzione della collezione

1. Informazioni generali sulla collezione
2. Specifica dei documenti da inserire
3. Configurazione
4. Costruzione
5. Visualizzazione della collezione

## Passi per la costruzione della collezione

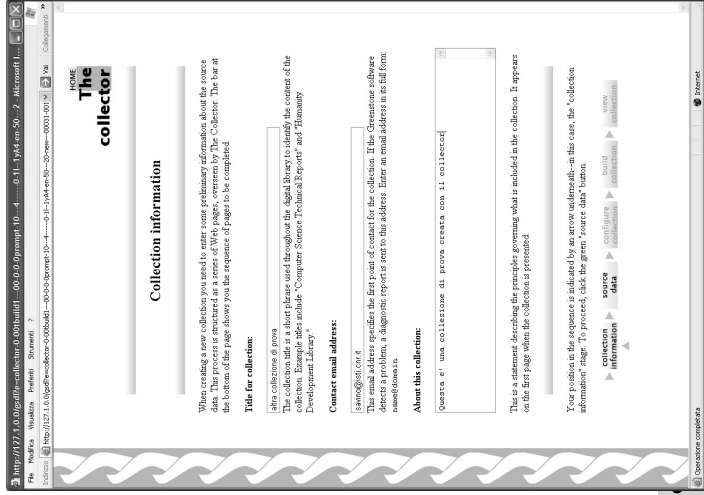
- Informazioni generali sulla collezione
- Specifica dei documenti da inserire
- Configurazione
- Costruzione
- Visualizzazione della collezione



## 1. Informazioni generali sulla collezione

- ◆ Specifica un nome per la collezione e fornisce informazioni associate alla collezione
  - Title
    - Breve frase utilizzata per identificare la collezione all'interno della Biblioteca Digitale
  - Contact e-mail address
  - Breve descrizione
    - Indica le caratteristiche della collezione ed i principi che definiscono cosa è incluso nella collezione

## Inserimento informazioni sulla collezione



## Passi per la costruzione della collezione

- Informazioni generali sulla collezione
- Specifica dei documenti da inserire
- Configurazione
- Costruzione
- Visualizzazione della collezione



## 2. Documenti da inserire nella collezione

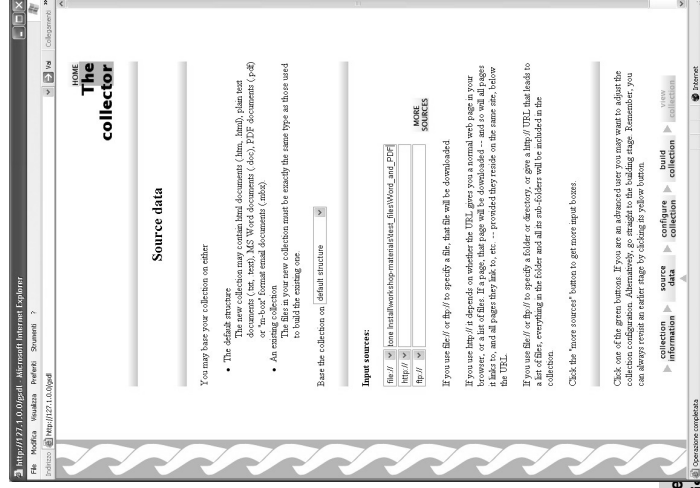
- ◆ Specifica dove si trovano i documenti da inserire
- ◆ Specifica delle sorgenti
  - *file://*
  - *http://*
  - *ftp://*
- ◆ File name sul Greenstone server system
  - Il file sarà incluso nella collezione
- ◆ Directory name sul server Greenstone
  - Tutti i documenti che si trovano nella directory e tutte le sottodirectory verranno incluse

*file://*

**http://**

- ◆ **Web page**
  - La web page verrà scaricata
  - Tutte le pagine alle quali la pagina punta e che si trovano sullo stesso sito verranno scaricate
- ◆ **URL che punta ad una lista di files**
  - Tutto quello che si trova nel folder e nei suoi sotto-folder sarà incluso nella collezione

## Dati da inserire nella collezione



**ftp://**

- ◆ **File verrà scaricato usando FTP**
- ◆ **Directory name sul server FTP**
  - Tutto quello che si trova nel folder e nei suoi sotto-folder sarà incluso nella collezione

## Passi per la costruzione della collezione

- Informazioni generali sulla collezione
- Specifica dei documenti da inserire
- Configurazione
- Costruzione
- Visualizzazione della collezione



### 3. Configurazione

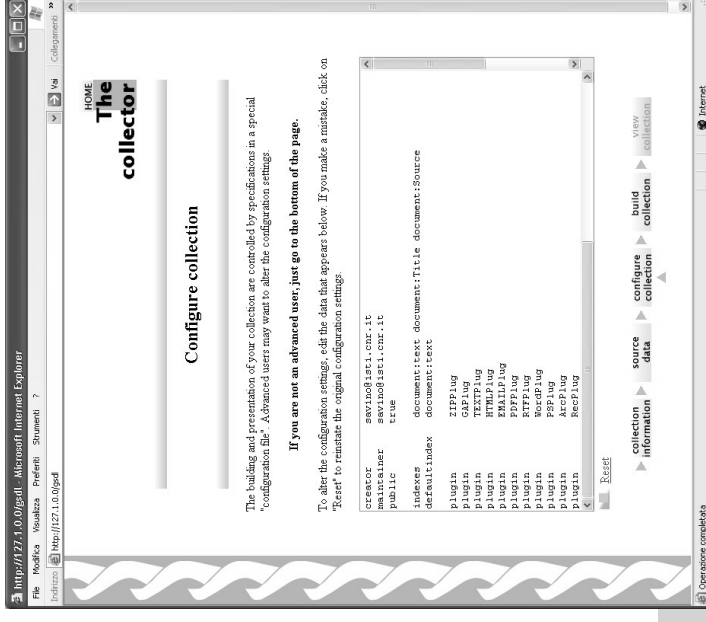
- ◆ Questo passo non verrà trattato
- ◆ Permette di adattare le opzioni di configurazione della collezione.
- ◆ La costruzione e la presentazione di tutte le collezioni sono controllate da specifiche che si trovano in un file di configurazione

### Passi per la costruzione della collezione

- Informazioni generali sulla collezione
- Specifica dei documenti da inserire
- Configurazione
- Costruzione
- Visualizzazione della collezione

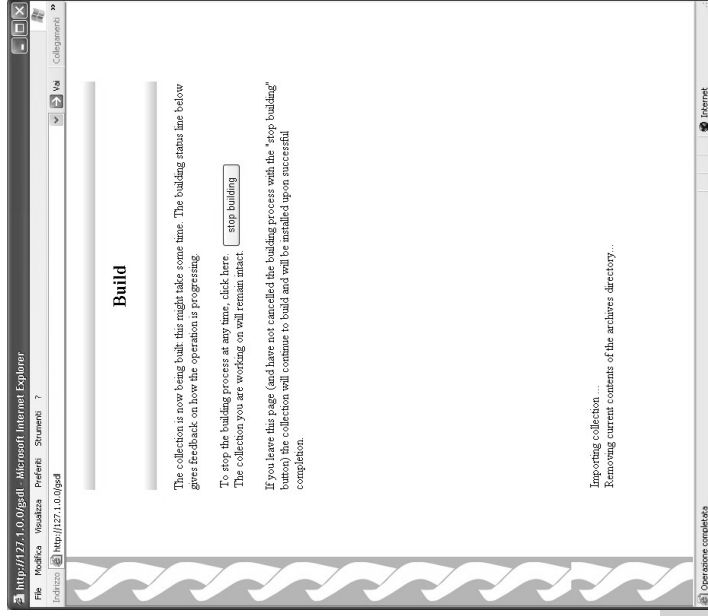


### Modifica della configurazione



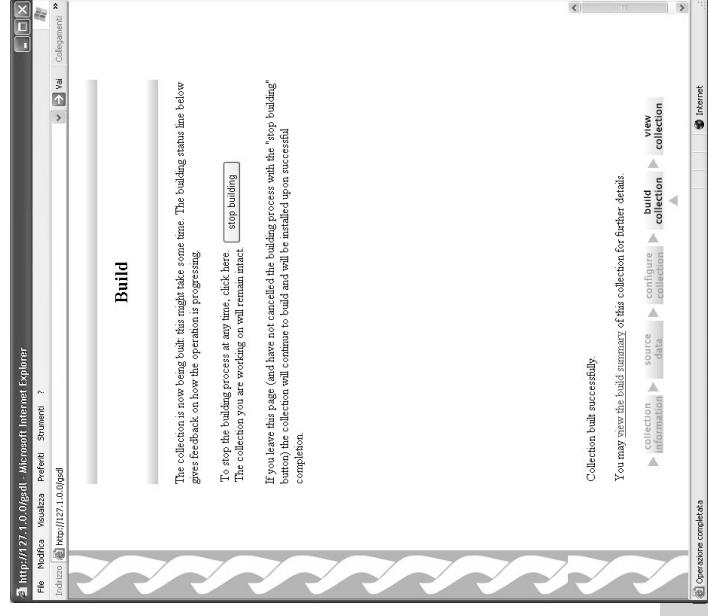
### 4. Costruzione

- ◆ La fase di costruzione della collezione viene eseguita autonomamente dal sistema
- ◆ Durante questa fase vengono costruiti gli indici:
  - Per il browsing
  - Per il searching
  - Vengono seguite le specifiche presenti nel file di configurazione
- ◆ Status line mostra come progredisce l'operazione



## Costruzione della collezione

Pasquale Savino – ISTI-CNR



## Costruzione della collezione

Pasquale Savino – ISTI-CNR

## Passi per la costruzione della collezione

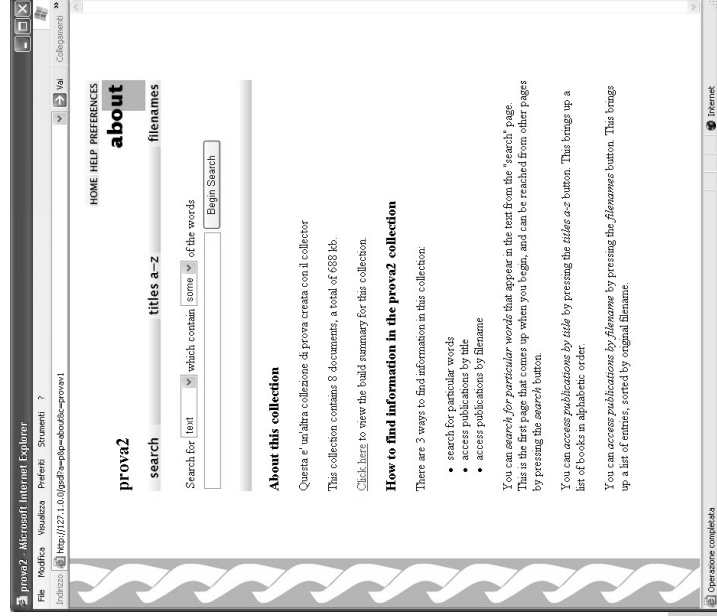
- Informazioni generali sulla collezione
- Specifica dei documenti da inserire
- Configurazione
- Costruzione
- Visualizzazione della collezione



Pasquale Savino – ISTI-CNR

Biblioteche Digitali  
Esempi – Il sistema Greenstone

63



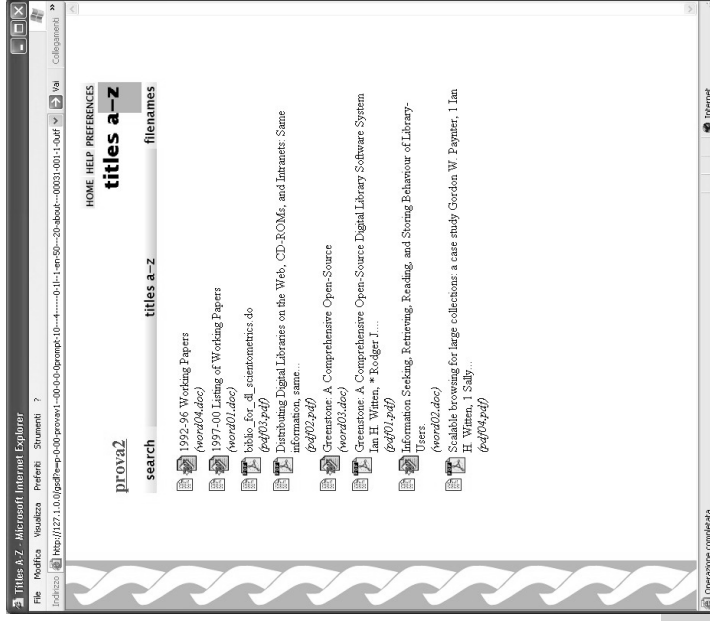
## 5. Visualizzazione

◆ Al termine della costruzione della collezione, questa può essere immediatamente visualizzata

Pasquale Savino – ISTI-CNR

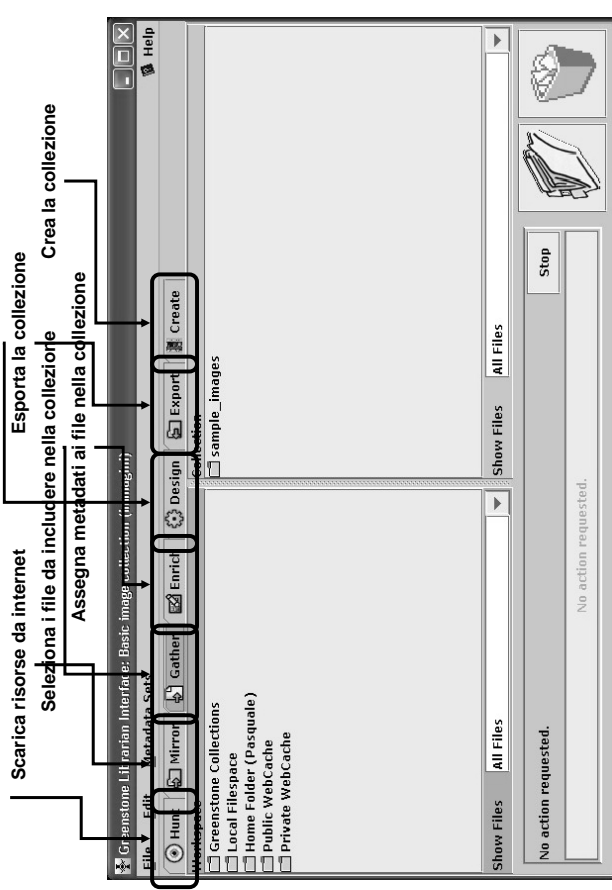


## Lista dei documenti inseriti nella nuova collezione



## The Greenstone Librarian Interface

### Localizza risorse da internet



Scarica risorse da internet  
Seleziona i file da includere nella collezione  
Assegna metadati ai file nella collezione

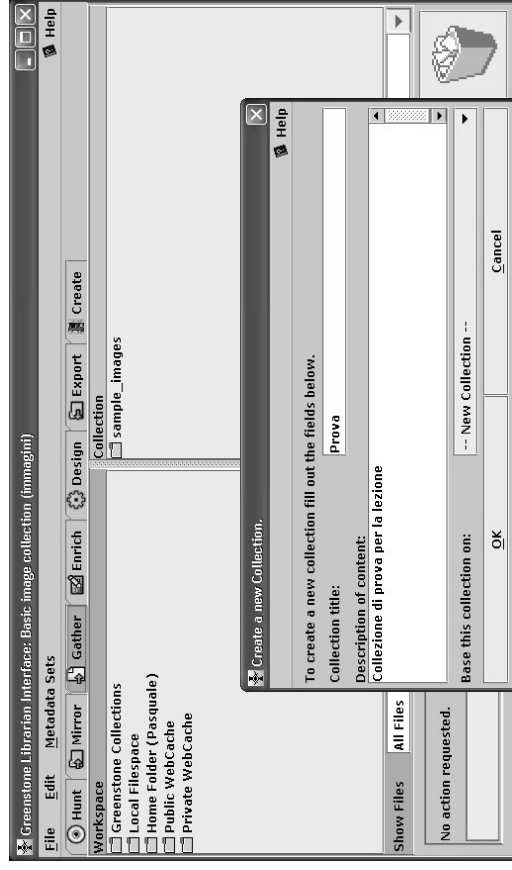
### Definisce le funzionalità e le modalità di presentazione della collezione

Esporta la collezione

Crea la collezione

Assegna metadati ai file nella collezione

## Crea una nuova collezione

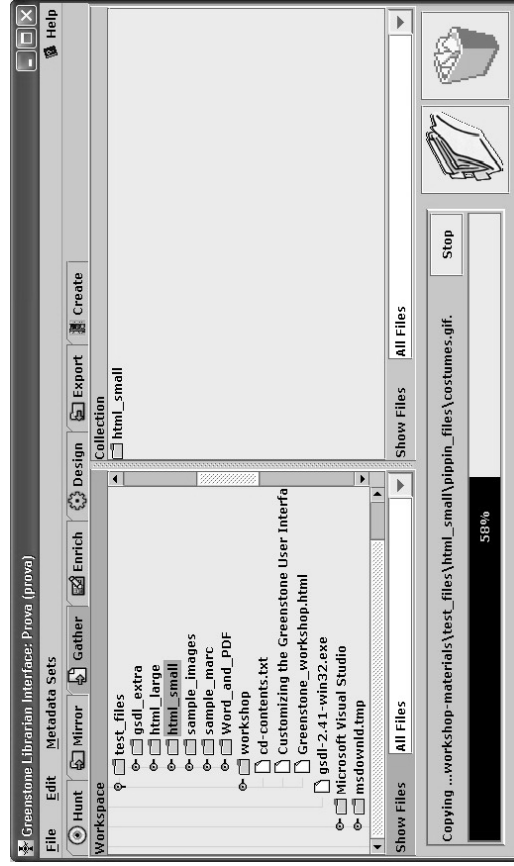


## Seleziona un set di metadati

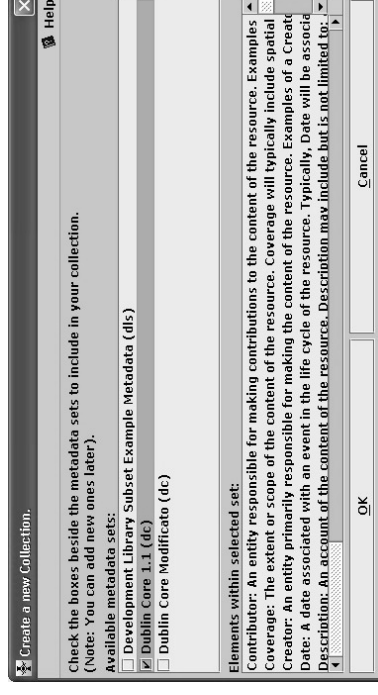
- ◆ Due insiemi predefiniti
  - Development Library Subset Example
  - Dublin Core 1.1



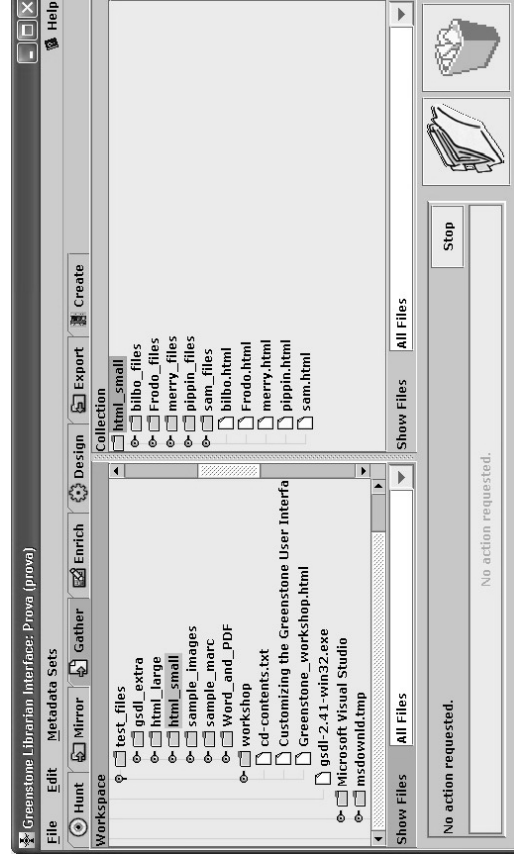
## Seleziono e copio dei file nella collezione



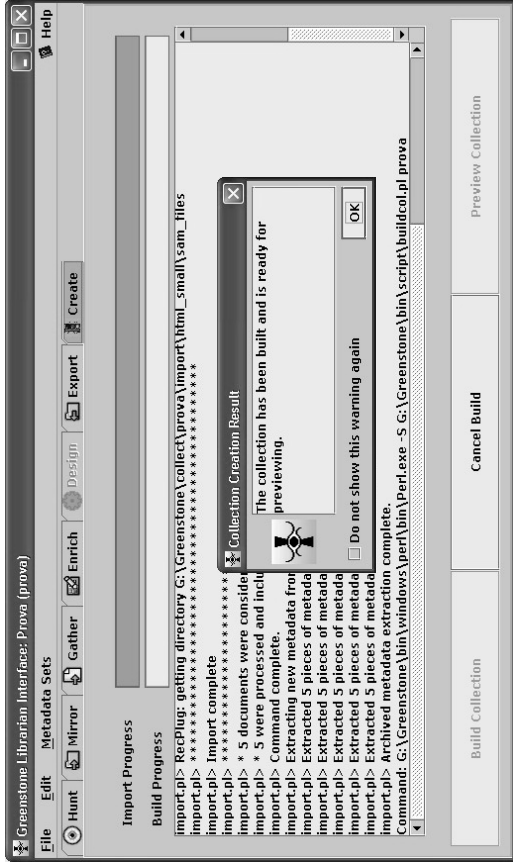
## Seleziono Dublin Core



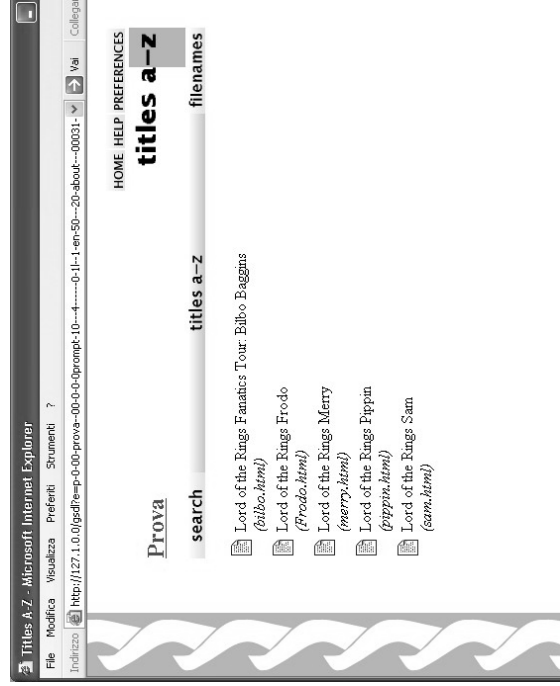
## File copiati nel panel di destra



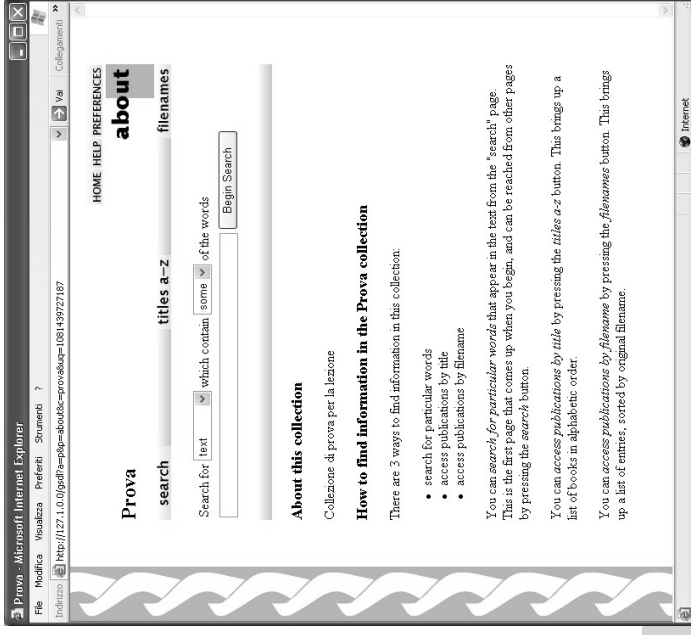
# Creo la collezione



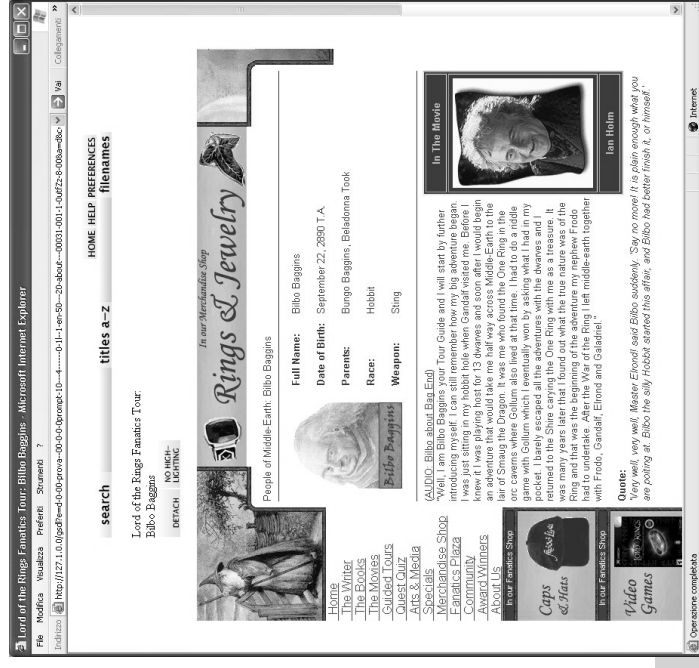
# Titoli in ordine alfabetico



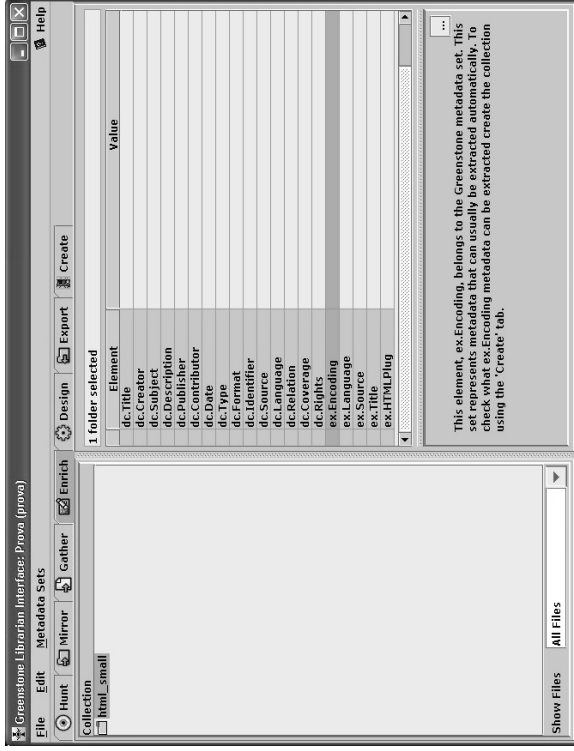
# Pagina Iniziale della Biblioteca Digitale



# Uno dei documenti della collezione



## Il modulo per la creazione dei metadati

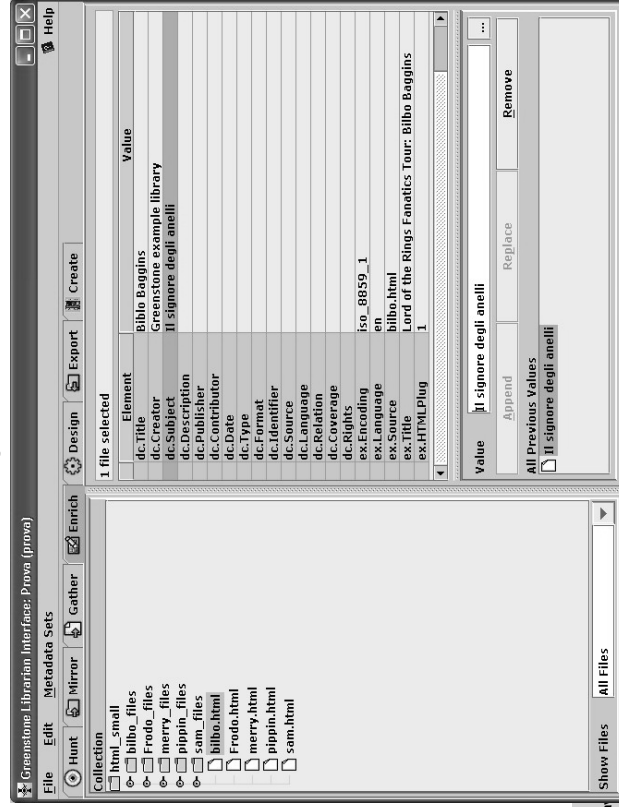


Pasquale Savino – ISTI-CNR

Biblioteche Digitali  
Esempi – Il sistema Greenstone

77

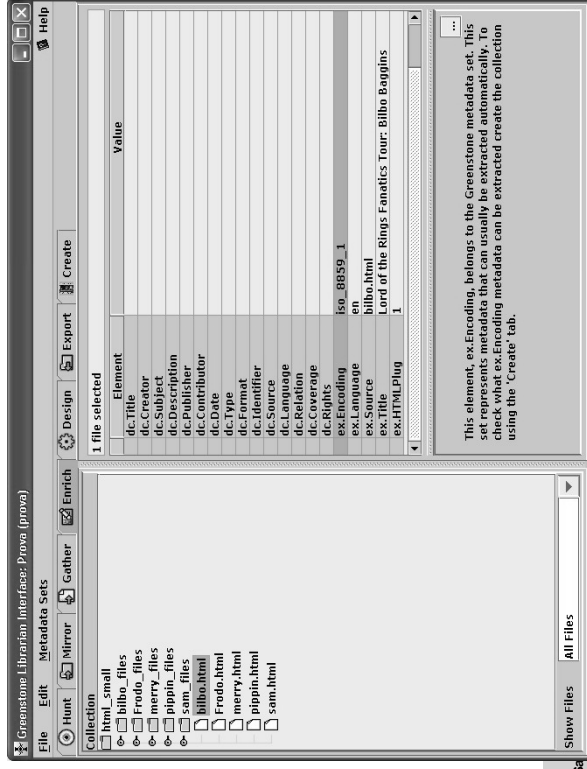
## Metadati assegnati manualmente



Pasquale Savino

79

## Metadati estratti automaticamente



Pasquale Savino

78