

## 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

Greenstone

7

### Esempi di sistemi Greenstone

Pasquale Savino – ISTI-CNR

## 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

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

7

### Esempi di sistemi Greenstone

Pasquale Savino – ISTI-CNR

Biblioteche Digitali

2

2

## 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

## Biblioteche Digitali che usano Greenstone

- Alcuni esempi
    - Archives of Indian Labour
      - <http://www.indiabolearchives.org/>
    - Chophin Early Editions
      - <http://cophilib.uchicago.edu/>
    - Lehigh University Digital Bridges Collection
      - <http://bridges.lib.lehigh.edu/index.html>
    - Mirabilia. Vincenzoferrati
      - <http://www.mirabilavincenzoferrati.org/sezioni/006/index.html>
    - New York Botanical Garden
      - <http://www.nybg.org/bcsl/tarin/page1.html>
    - Project Gutenberg
      - <http://public.library.org/gutenberg/cgi-bin/library.cgi?a=p&p=about&c=gberg>
    - State Library of Tasmania. Sheet Music Collection
      - <http://greenstone.statelibrary.tas.gov.au/>
    - Washington Research Library Consortium Special Collections
      - <http://www.adain.wrlc.org/all/>

# University of Chicago Library

The University of Chicago  
LIBRARY

## ▪ Chopin Early Editions

Chopin Early Editions (1833-1835)  
titles  
uniform titles  
titles  
uniform titles  
titles  
titles  
titles

The University of Chicago  
LIBRARY

HELP REFERENCES

dedications

genres

Page 2

[view score](#) [bibliographic description](#)

[Page 2] [\[Go\]](#)

▼

view score [bibliographic description](#)

[Page 2] [\[Go\]](#)

▼

Uniform Title: Recital piano op. 1. C major.  
Title: Adieu à Varsovie; Consolation; C minor.

Composer: Chopin, Frédéric, 1810-1849.

Conductor: Tafelmusik Baroque Ensemble.

Publisher: Hinrichsen, Leipzig, 1849.

Publisher: Hinrichsen, Leipzig, 1849.

Co-publisher: Kistner & Co., New York.

Co-publisher: Hinrichsen, Leipzig, 1849.

Description: 13 p. : piano.

Notes: Composed in 1830-1831.

Quoted in: "Frédéric Chopin," in: "Musical Times," 1875, p. 111.

Also quoted in: "Frédéric Chopin," in: "Musical Times," 1875, p. 111.

Also quoted in: "Frédéric Chopin," in: "Musical Times," 1875, p. 111.

Also quoted in: "Frédéric Chopin," in: "Musical Times," 1875, p. 111.

Also quoted in: "Frédéric Chopin," in: "Musical Times," 1875, p. 111.

Also quoted in: "Frédéric Chopin," in: "Musical Times," 1875, p. 111.

Also quoted in: "Frédéric Chopin," in: "Musical Times," 1875, p. 111.

Also quoted in: "Frédéric Chopin," in: "Musical Times," 1875, p. 111.

Also quoted in: "Frédéric Chopin," in: "Musical Times," 1875, p. 111.

Also quoted in: "Frédéric Chopin," in: "Musical Times," 1875, p. 111.

Also quoted in: "Frédéric Chopin," in: "Musical Times," 1875, p. 111.

Also quoted in: "Frédéric Chopin," in: "Musical Times," 1875, p. 111.

Also quoted in: "Frédéric Chopin," in: "Musical Times," 1875, p. 111.

Also quoted in: "Frédéric Chopin," in: "Musical Times," 1875, p. 111.

Also quoted in: "Frédéric Chopin," in: "Musical Times," 1875, p. 111.

Also quoted in: "Frédéric Chopin," in: "Musical Times," 1875, p. 111.

Also quoted in: "Frédéric Chopin," in: "Musical Times," 1875, p. 111.

Also quoted in: "Frédéric Chopin," in: "Musical Times," 1875, p. 111.

Also quoted in: "Frédéric Chopin," in: "Musical Times," 1875, p. 111.

Also quoted in: "Frédéric Chopin," in: "Musical Times," 1875, p. 111.

Also quoted in: "Frédéric Chopin," in: "Musical Times," 1875, p. 111.

Also quoted in: "Frédéric Chopin," in: "Musical Times," 1875, p. 111.

Also quoted in: "Frédéric Chopin," in: "Musical Times," 1875, p. 111.

Also quoted in: "Frédéric Chopin," in: "Musical Times," 1875, p. 111.

Also quoted in: "Frédéric Chopin," in: "Musical Times," 1875, p. 111.

Also quoted in: "Frédéric Chopin," in: "Musical Times," 1875, p. 111.

Also quoted in: "Frédéric Chopin," in: "Musical Times," 1875, p. 111.

Also quoted in: "Frédéric Chopin," in: "Musical Times," 1875, p. 111.

Subjects: Recital (Piano)

Piano music.

Plate No.: 1423

Plate No.: 1423

Call No.: M25-54823 t.1

Call No.: M25-54823 t.1

## 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

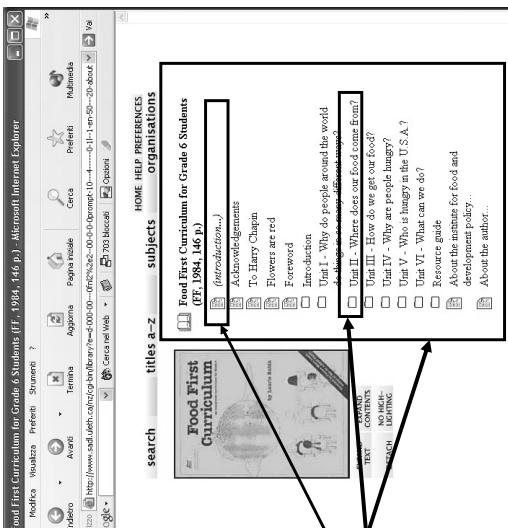
Biblioteche Digitali  
Esempi – Il sistema Greenstone

9

Biblioteche Digitali  
Esempi – Il sistema Greenstone

10

## Modello documento gerarchico



I metadati possono essere associati ad ogni livello del documento

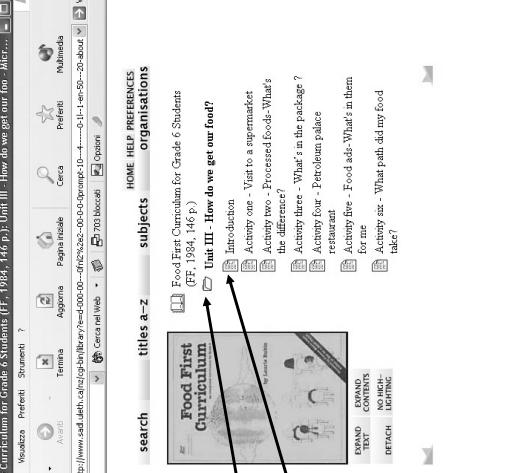
## Testo strutturato gerarchicamente

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

Biblioteche Digitali  
Esempi – Il sistema Greenstone

10

## Modello documento gerarchico



Operatore operatore...3

Biblioteche Digitali  
Esempi – Il sistema Greenstone

12

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.biblio.org/gutenberg/library.cgi?a=p&p=about&c=gberg>

◆ **Esempio: Savino – ISTI-CNR**

"You know it was supposed that old Sir Jasper, being a bachelor, would leave his fortune to his two nephews. But he was an oddity, and as the title must go to young Jasper by right, the old man said Maurice should have the money. He was poor, young Jasper rich, and it seemed but just, though Madame Marie was very angry when she learned how the will was made."

"But Maurice didn't get the fortune. How was that?"

See  [ISTI-CNR](#)

Pasquale Savino – ISTI-CNR

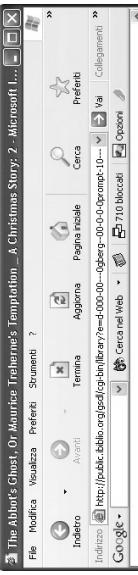
13

Biblioteche Digitali  Esempi – Il sistema Greenstone

Ascolese Savino – ISTI-CNR

Internet 

Testo non  
strutturato



Immagini delle pagine e Testo estratto

- ◆ **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**
  - ◆ **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.sad.uleth.ca/nzcg-bin/miuepsalibRARY?e=d-0niuapea-00-0-0-010-4---full---0-11--1t-50---20-about---00001-001-outfzz-8-a=p&p=about>

**Biblioteche Digitali**  
**Esempi - Il sistema Greenstone**

## Altri tipi di documenti

- ◆ **Audio ed immagini fotografiche**
    - Esempio: Oral History Collection. Mirabilis Vicomercati
      - <http://www.sadl.uieht.ca/nz/cgi-bin/library?a=p&p=about&c=oh>
      - <http://www.mirabilisvicomercati.org/gsdl/cgi-bin/library?a=p&p=about&c=mu>
  - ◆ **Video**
    - Esempio: Music Video Collection
      - <http://www.sadl.uieht.ca/nz/cgi-bin/library?a=p&p=about&c=mu>
  - ◆ **Musica**
    - Vari rappresentazioni: spartiti, MIDI, musica sintetizzata, ecc.
      - Esempio: Music Digital Library
        - <http://www.sadl.uieht.ca/nz/cgi-bin/musiclibrary>
  - ◆ **Lingue multiple**
    - Interfaccia e documenti
      - Esempio: Arabic Collection
        - <http://www.sadl.uieht.ca/nz/cgi-bin/library?a=p&p=about&c=arab>

Immagini

A screenshot of Microsoft Internet Explorer displaying a search results page. The search term 'titoli' is entered in the address bar. The results page shows a grid of nine images, each labeled with a caption: 'Abitazione della casa a corte di Via San Gerolamo', 'Abitazione della casa a corte di Via San Gerolamo', 'Abitazione della casa a corte di Via San Gerolamo', 'Abside della Chiesa di Santo Stefano', 'Abside della Chiesa di Santo Stefano', 'Abside della Chiesa di Santo Stefano', and 'Abside della Chiesa di Santo Stefano'. The browser interface includes a toolbar at the top with icons for Back, Forward, Stop, Refresh, Home, and Favorites. The address bar shows the URL 'http://www.mirabilis.vicomercati.org/gestione/bsn/bsn?m=000-00-00-G-0-0-pronc>10'. The title bar reads 'Mirabilis Vicomercati - Archivi - Archivio Fotografico - Microsoft Internet Explorer'. The left sidebar contains links for 'Mirabilis Vicomercati', 'Archivi', 'Archivio Fotografico', and 'Archivio'. The right sidebar has links for 'PAGINA PRINCIPALE AUTI IMPOSTAZIONI', 'titoli', 'titoli', 'autori', 'soggetti', 'persone', 'date', and 'cerca'.

Pasquale Savino – ISTI-CNR

Audio

Friday night was a big night in Hamilton - The Golden Mile during the 1940s. Memories of... [View](#)

**File** **Media** **Version** **Preprint** **Strumenti** ?

[Dettagli](#) [https://www.semanticscience.org/bibliography?eid=60000000-0000-0000-0000-000000000000&id=10-4-4](#) [Via](#)

**search** **titles a-z** **authors a-z**

**NO LIGHTING** **DETACH TEXT** **EXPAND**

"Friday night was a big night in Hamilton": the Golden Mile during the 1940s. Memories of... [View](#)

Memories of WWII [View](#)

Remembers Patriotic Hut in Garden Place, patriotic parades, end of the war. [View](#)

"That was a day everyone went absolutely mad, all the teachers went mad and down from the school, down everybody sort of狂妄化 or狂想曲 or something and right down into town, everybody got lost. I think it was one day we were allowed just to do as we liked!"

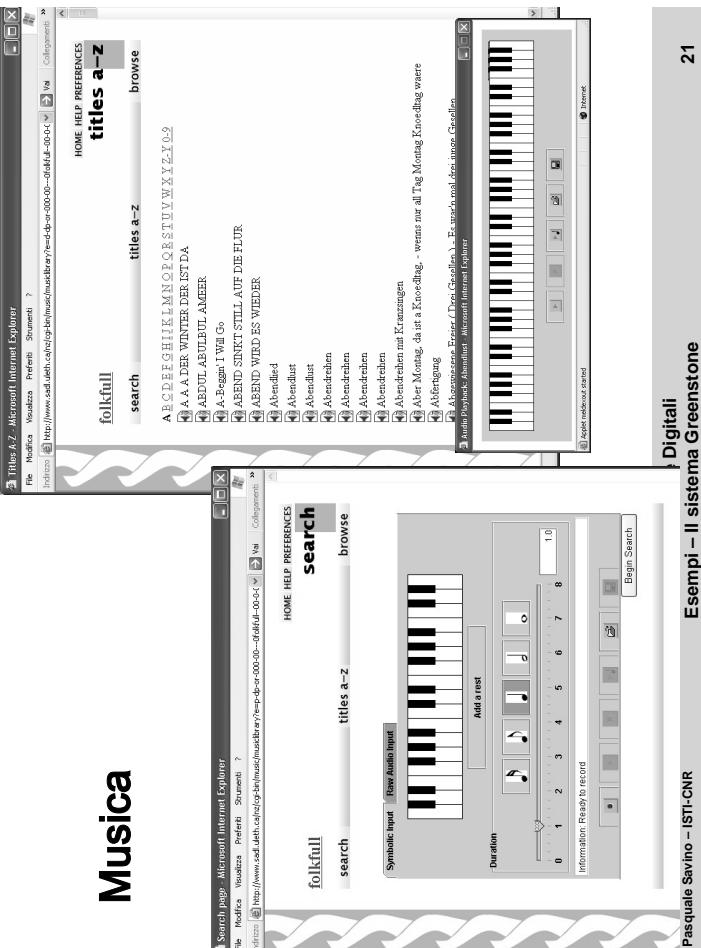
**HOME** **HELP** **PREFERENCES** **authors a-z**

**13**

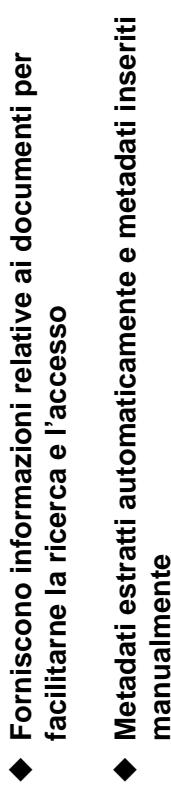
**II sistemi Greenstone**

Pasquale Savino – ISTI-CNR

Musica



## Metadata



## Esempi di metadati

- ◀ **Metadata bibliografici convenzionali**
  - Title
  - Author
  - Date
  - Publication
  - Volume Number
  - Issue Number
  - Page Numbers
  - ▶ **MARC**
  - ▶ **Dublin Core**

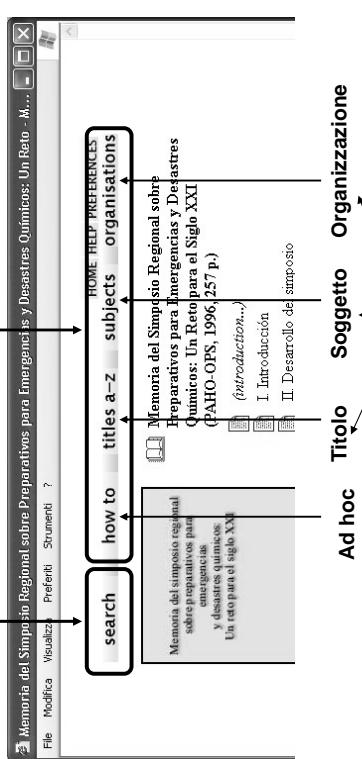
## Sommaio

- ◆ 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 e browsing

## Searching

Browsing basato sull'uso dei metadati | Searching



## ◆ Tipi di interrogazioni (query) in Greenstone

- ◆ Case-folding e stemming
  - ◆ Ricerca di frasi
  - ◆ Diverse interfacce di interrogazione



卷之三

27

Biblioteca Digitale

Biblioteca Digital

20

## Tipi di interrogazioni

## ◆ 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

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.

Provare lo stesso query "Italy :Environment"

## 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**

Biggest movie ever made

20

27

Biblioteca Digitale

Biblioteche Digitali

20

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

27

Biblioteche Digitali  
Esempi - Il sistema Greenstone

27

## Ricerca di frasi

- ◆ **Ricerca di un gruppo di parole contigue**
  - ◆ **Due diverse modalità di ricerca di frasi:**
    - Postretrieval scan
      - Ricerca i documenti che contengono i termini inseriti
      - Determina se i termini sono contenuti in documenti ritrovati
      - Richiede lo stesso indice usato per le singole, ma le ricerche possono essere eseguite contemporaneamente
    - Word-level index
      - Viene effettuata la ricerca della frase specifico che contiene informazioni all'interno di ogni documento
      - Questi indici permettono di determinare se due termini sono consecutivi senza accedere al database
      - Gli indici sono più grandi di quelli singoli termini, ma la ricerca di frasi è più veloce

Pasquale Saviglio - ISTI-CNR

29  
Environ Monit Assess (2007) 130:29–36

Pascuale Savino - ISTI-CNR | Ecomai - sistema Gransteno

Browsing



HOME HELP PREFERENCES

Browsing

- ## ◆ Browsing per data

dates  
major newspapers

māori newspapers	dates	series	dates
search			
1842-1843	1844-1846	1848-1856	1857-1858
1853-1882	1883-1894	1895-1896	1897-1898
1909-1910	1911-1912	1913-1914	1915-1916
1917-1918	1919-1921	1921-1924	1924-1925
1842 January	[icon]	Te Kākete o Nu Tētē Volume 1	January 1, 1842
		Number 1	
February	[icon]	Te Kākete o Nu Tētē Volume 1	February 1, 1842
		Number 2	
March	[icon]	Te Kākete o Nu Tētē Volume 1	March 1, 1842
		Number 3	
April	[icon]	Te Kākete o Nu Tētē Volume 1	April 1, 1842
		Number 4	
May	[icon]	Te Kākete o Nu Tētē Volume 1	May 2, 1842
		Number 5	
June	[icon]	Te Kākete o Nu Tētē Volume 1	June 1, 1842
		Number 6	
July	[icon]	Te Kākete o Nu Tētē Volume 1	July 1, 1842
		Number 7	
August	[icon]	Te Kākete o Nu Tētē Volume 1	August 1, 1842

## Diverse interfacce di interrogazione

## Browsing

- ◆ **Browsing basato su strutture gerarchiche di classificazione**
- ◆ **Ad es. in figura viene mostrato il primo livello di subject**
- ◆ **Nelle due figure accanto vengono mostrati i due livelli successivi per subject**

The screenshot shows a hierarchical classification tree. The root node is "01.00 General reference". Below it are several categories: "02.00 Sustainable Development, International cooperation, Projects, NGO, Organizations, Poverty and Hunger Alleviation, Basic Human Needs", "03.00 Economics, Economic policies, Finance, Trade, Structural Adjustment, Macroeconomics, Cooperatives", "04.00 Business, Multi-enterprises, Management, Marketing, Finance", "05.00 Politics, Public administration, Law and Legislation, Peace, Human Rights", "06.00 Society, Culture, Conservancy, Women, Youth, Population", "07.00 Education and Training", "08.00 Communication, Information and Documentation", "09.00 Transport", "10.00 Science, Research and Technology", "11.00 Ecology, Bio-diversity, Natural resources, Environment, Waste Disposal", "12.00 Water, Sanitation and Sewage Disposal", "13.00 Health, Nutrition, Medicine", "14.00 Agriculture and Food Processing", "15.00 Plant and Crop Production, Protection, and Processing", "16.00 Animal Husbandry and Animal Product Processing", "17.00 Forestry and Aquaculture", and "18.00 Forestry and Agroforestry".

Pasquale Savino – ISTI-CNR

Esempi – Il sistema Greenstone

## Browsing

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

The screenshot shows a hierarchical classification tree. The root node is "01.00 General reference". Below it are several categories: "02.00 Sustainable Development, International cooperation, Projects, NGO, Organizations, Poverty and Hunger Alleviation, Basic Human Needs", "03.00 Economics, Economic policies, Finance, Trade, Structural Adjustment, Macroeconomics, Cooperatives", "04.00 Business, Multi-enterprises, Management, Marketing, Finance", "05.00 Politics, Public administration, Law and Legislation, Peace, Human Rights", "06.00 Society, Culture, Conservancy, Women, Youth, Population", "07.00 Education and Training", "08.00 Communication, Information and Documentation", "09.00 Transport", "10.00 Science, Research and Technology", "11.00 Ecology, Bio-diversity, Natural resources, Environment, Waste Disposal", "12.00 Water, Sanitation and Sewage Disposal", "13.00 Health, Nutrition, Medicine", "14.00 Agriculture and Food Processing", "15.00 Plant and Crop Production, Protection, and Processing", "16.00 Animal Husbandry and Animal Product Processing", "17.00 Forestry and Aquaculture", and "18.00 Forestry and Agroforestry".

Pasquale Savino – ISTI-CNR

Biblioteca Esempi – Il siste...

## Browsing su frasi

- ◆ Associate metadati strutturati per la classificazione può essere molto oneroso.
- ◆ Se questa informazione non è disponibile, in Grenstone 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

Search frasi che

contengono la

parola "locust"

Vengono visualizzate le frasi con informazioni sul numero di documenti

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

The screenshot shows a search interface for the "phind classifier demo". The search term is "locust". The results table shows the following data:

phrase	docs	freq
locust	719	1767
locust numbers	127	285
migratory locust	89	132
locust infestation	84	112
locust	64	110
locust satiation	66	94
locust damage	50	85
locust control	45	84
forest-est locust	46	76
Get more phrases	25	59

Pasquale Savino – ISTI-CNR

Biblioteche Digitali

Esempi – Il sistema Greenstone

Esempi – Il Document Drive

Pasquale Savino – ISTI-CNR

Quindi posso visualizzare il documento

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

## 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

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

Il "Collector" di Greenstone

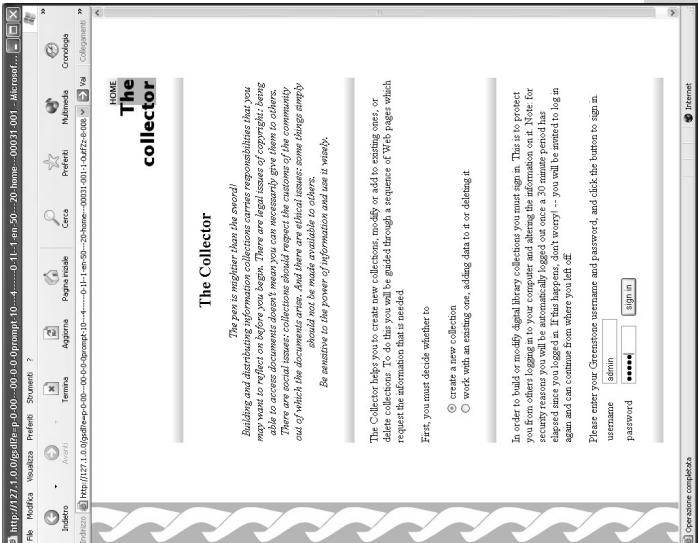
- ◆ **Il Collector permette di**
    - Creare una nuova collezione di documenti
    - Modificare o aggiungere documenti ad una collezione esistente
    - Cancellare una collezione

Biblioteche Digitali

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

**La home  
page del  
collector**



Biblioteche Digitali

Biblioteche Digitali

卷之三

## ◆ Uso della Librarian Interface

- **Uso della Librarian Interface**
  - Funziona in locale
  - Interfaccia per creare nuovi valori di metadati

erface

- **Uso della Librarian Interface**
  - Funziona in locale
  - Interfaccia per creare nuovi valori di metadati

Biblioteche Digitali

卷之三

Pasquale Savino – ISTI-CNR 43

Ese Internet Operazione complesso

Pasquale Savino - ISTI-CNR



# Passi per la costruzione della collezione

**Inserimento informazioni sulla collezione**

**Collection information**

**Title for collection:** The collector

**Contact email address:** santo@isrc.it

**About this collection:** Questa e' una collezione di prova creata con 11 collattate.

**Summary:** When creating a new collection you need to enter some preliminary information about the source data. This process is illustrated in a series of Web pages overseen by The Collector. The bar at the bottom of the page shows you the sequence of pages to be completed.

**Title for collection:** The collector

**Description & goals:** This collection is a short thesis used throughout the legal library to identify the contents of the collection. Entries may include Computer Science Technical Reports and Theses.

**Development Library:** The collector is used throughout the legal library to identify the contents of the collection. Entries may include Computer Science Technical Reports and Theses.

**Source:** santo@isrc.it

**Information:** The email address specifies the electronic address for the collection. If the Greenstone software detects a problem, a diagnostic message is sent to this address. Enter an email address in its full form, including domain name.

**About this collection:** Questa e' una collezione di prova creata con 11 collattate.

**Summary:** This is a statement describing the principles governing what is included in the collection. It appears on the first page whenever the collection is presented.

**Source:** You're position in the sequence is indicated by an arrow underneath—in this case, the "collection information" page. To proceed, click the green "source data" button.

**Information:** collection ▶ source ▶ configuration ▶ build ▶ collections ▶ view ▶ collections ▶ collections

**Biblioteca Digitali Esempi – Il sistema Greenstone**

Pasquale Savino – ISTI-CNR

## 2. Documenti da inserire nella collezione

- ◆ Specifica dove si trovano i documenti da inserire
- ◆ Specifica delle sorgenti
  - file://
  - http://
  - ftp://

**Biblioteche Digitali Esempi – Il sistema Greenstone**

Pasquale Savino – ISTI-CNR

Esempi – Il sistema Greenstone

**file://**

- ◆ **File name sul server 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

**http://**

**ftp://**

◆ **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

Biblioteche Digitali  
Esempi – Il sistema Greenstone  
Pasquale Savino – ISTI-CNR  
53

The screenshot shows a web-based application titled "The collector". On the left, there's a sidebar with a tree view of a folder structure containing "index.html", "index.htm", "index.cgi", and "index.php". The main area displays a list of URLs under the heading "Input sources:":

- file:///home/fisical/workshop/materials/first/theWord\_and\_PDF/
- http://
- ftp://

Below the list, there are buttons for "MORE" and "SOURCES". To the right, there's a detailed description of how URLs are handled by the collector.

## Dati da inserire nella collezione

Check one of the green buttons. If you an advanced user you may want to adjust the collection configuration. Alternatively, go straight to the building page. Remember, you can always return an earlier stage by clicking the yellow 'button'.

If you use file:// or http:// to specify a file, then the file will be downloaded and it depends on whether the URL gives you a normal web page in your browser, or a list of files. In a page that page will be downloaded - and so will its pages if it has any, and all pages they link to, etc. - provided they reside on the same site as those used in the URL.

You can use file:// or http:// to specify a folder or directory of give a http://URL that leads to a list of files, everything in the folder and its sub-folders will be crawled in the collection.

Click the "more sources" button to get more input boxes.

Biblioteche Digitali  
Esempi – Il sistema Greenstone  
Pasquale Savino – ISTI-CNR  
53

Biblioteche Digitali  
Esempi – Il sistema Greenstone  
Pasquale Savino – ISTI-CNR  
54

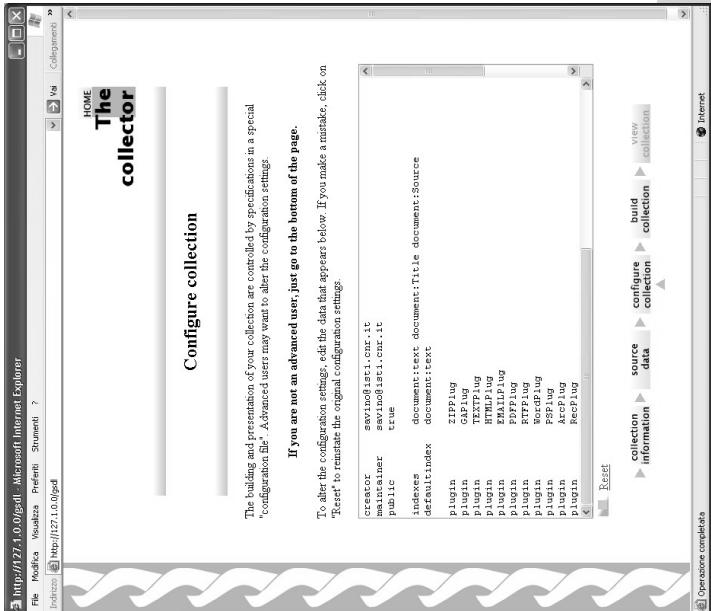
## Passi per la costruzione della collezione

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

Biblioteche Digitali  
Esempi – Il sistema Greenstone  
Pasquale Savino – ISTI-CNR  
56

## 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



## Modifica della configurazione

## Passi per la costruzione della collezione

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



## 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 progetisce l'operazione

## Costruzione della collezione

The collection is now being built; this might take some time. The building status line below gives feedback on how the operation is progressing.  
To stop the building process at any time, click here.   
The collection you are working on will remain intact.  
If you leave this page (and have not cancelled the building process with the "stop building" button) the collection will continue to build and will be installed upon successful completion.

Importing collection  
Removing current contents of the archive directory...

## Costruzione della collezione

Build

The collection is now being built; this might take some time. The building status line below gives feedback on how the operation is progressing.  
To stop the building process at any time, click here.   
The collection you are working on will remain intact.  
If you leave this page (and have not cancelled the building process with the "stop building" button) the collection will continue to build and will be installed upon successful completion.

Collection built successfully.  
You may [view the build summary](#) of this collection for further details.

## Passi per la costruzione della collezione

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



## 5. Visualizzazione

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

## Costruzione della collezione

Build

The collection is now being built; this might take some time. The building status line below gives feedback on how the operation is progressing.  
To stop the building process at any time, click here.   
The collection you are working on will remain intact.  
If you leave this page (and have not cancelled the building process with the "stop building" button) the collection will continue to build and will be installed upon successful completion.

Collection built successfully.  
You may [view the build summary](#) of this collection for further details.

provav2 Microsoft Internet Explorer  
File Modifica Visualizza Preferiti Strumenti ?  
Indirizzo: <http://127.0.0.1:0/gpds/provav2/index.html> Coleggimenti

HOME HELP PREFERENCES  
**about**

provav2 Microsoft Internet Explorer  
File Modifica Visualizza Preferiti Strumenti ?  
Indirizzo: <http://127.0.0.1:0/gpds/provav2/index.html> Coleggimenti

Quella è un'altra collezione di prova creata con il collector provav2.  
This collection contains 8 documents, a total of 688 kb.  
Click here to view the build summary for this collection.

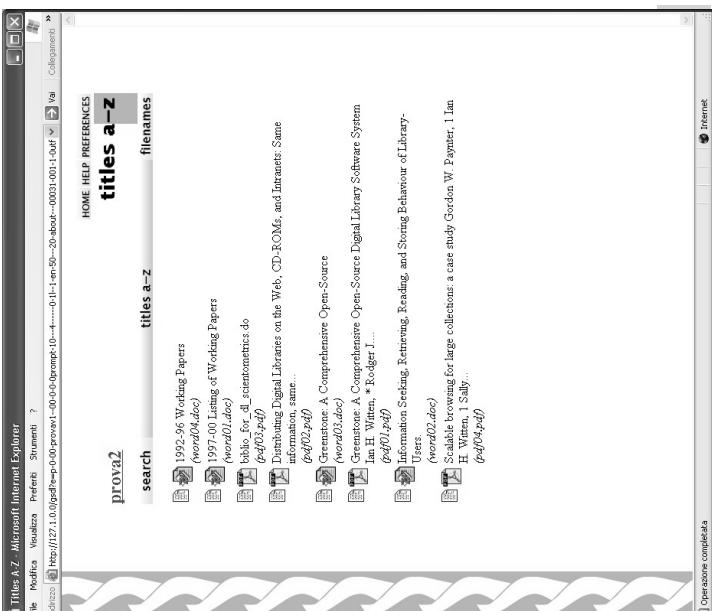
**About this collection**  
There are 3 ways to find information in this collection:

- search for particular words
- access publications by title
- access publications by filename

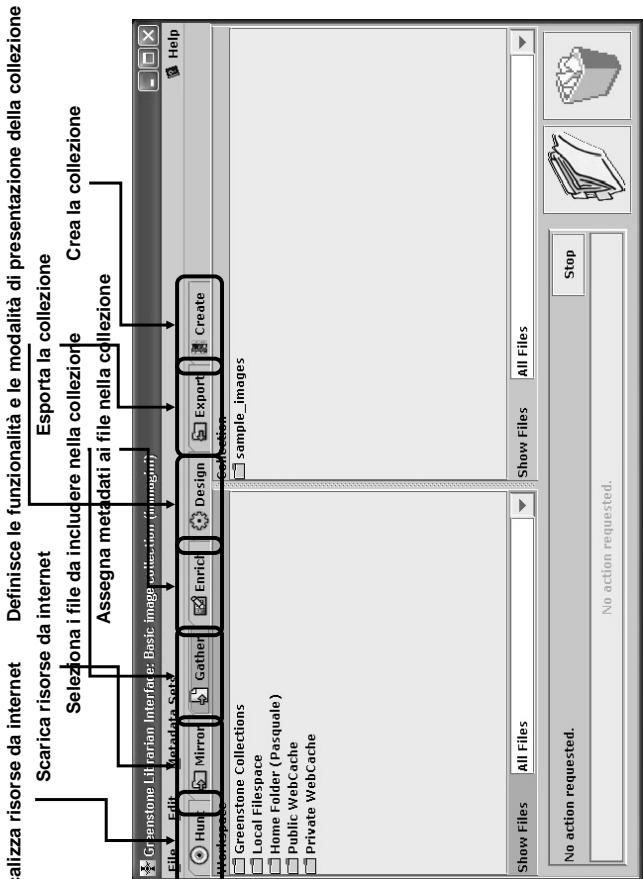
You can search for particular words that appear in the text from the "search" page. This is the first page that comes up when you begin, and can be reached from other pages by pressing the search button.  
You can access publications by title by pressing the titles a-z button. This brings up a list of books in alphabetical order.  
You can access publications by filename by pressing the filenames button. This brings up a list of entries, sorted by original filename.

**How to find information in the provav2 collection**

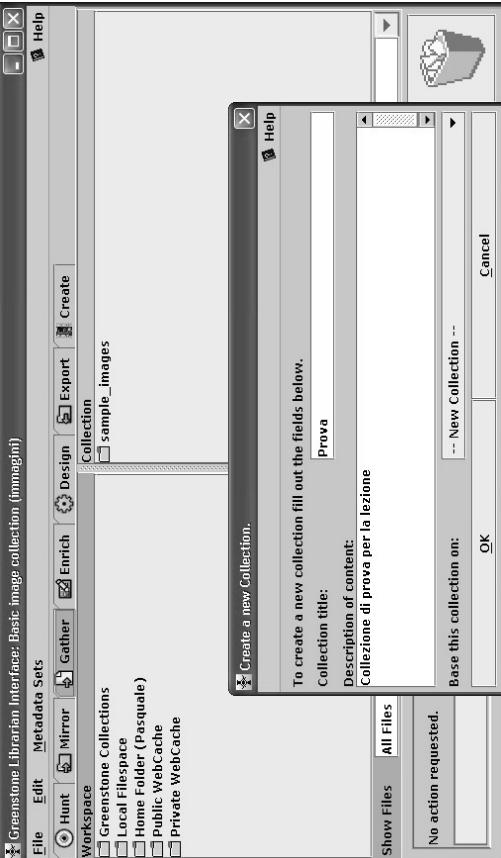
## **Lista dei documenti inseriti nella nuova collezione**



The Greenstone Librarian Interface



**Crea una nuova collezione**



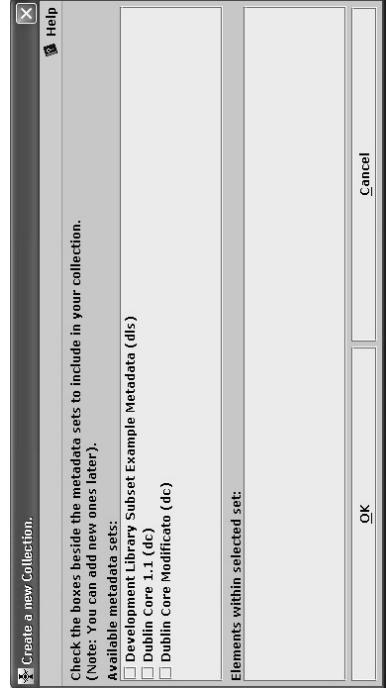
Pasquale Savino – ISTI-CNR      Biblioteche Digitali      Esempi – Il sistema Greenstone      6

**Biblioteche Digitali**  
**Esempi – Il sistema Greenstone**



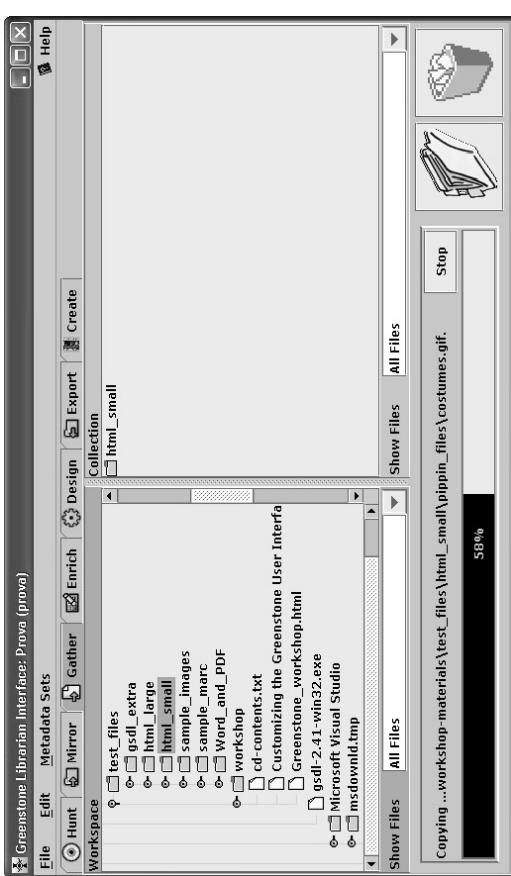
## Selezione un set di metadati

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



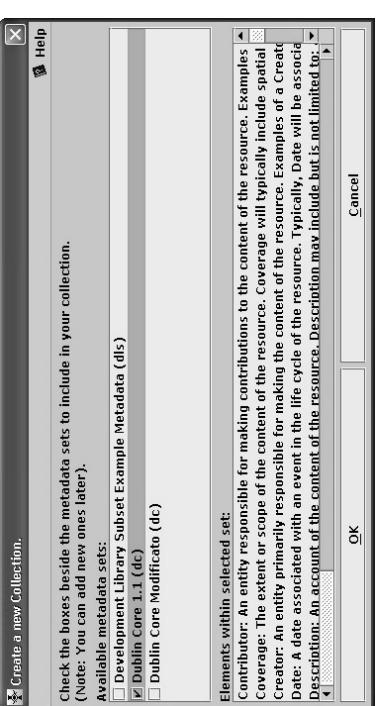
Pasquale Savino – ISTI-CNR      Biblioteche Digitali      Esempi – Il sistema Greenstone      69

## Selezione e copio dei file nella collezione



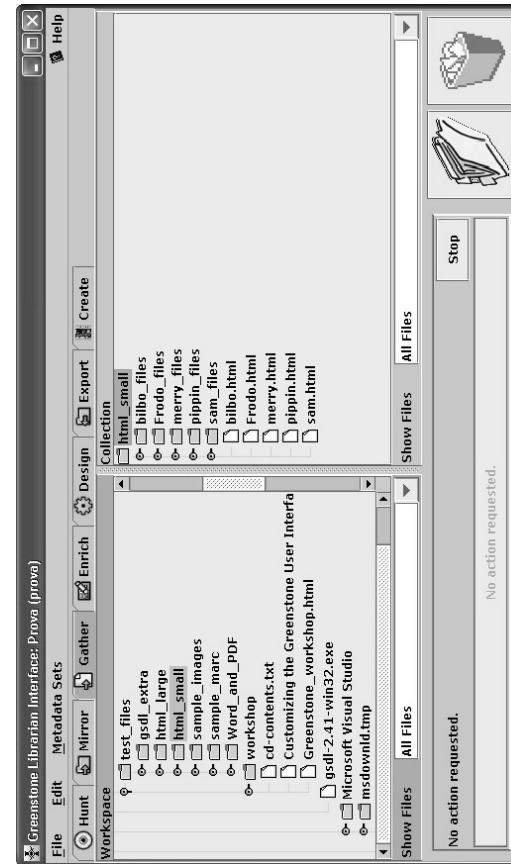
Pasquale Savino – ISTI-CNR      Biblioteche Digitali      Esempi – Il sistema Greenstone      71

## Selezione Dublin Core



Pasquale Savino – ISTI-CNR      Biblioteche Digitali      Esempi – Il sistema Greenstone      70

## File copiati nel panel di destra



Biblioteche Digitali  
Esempi – Il sistema Greenstone

Creo la collezione

Pagina  
Iniziale  
della  
Biblioteca  
Digitale

**Biblioteche Digitali**  
**Esempi – Il sistema Greenstone**

The screenshot shows the Prova library catalog search interface. At the top, there are two tabs: "filenames" (selected) and "titles a-z". Below the tabs is a search bar with the placeholder "Search for test". To the right of the search bar are three dropdown menus: "which contain", "some of the words", and "all of the words". Further to the right are two buttons: "Begin Search" and "End Search".

**About this collection**

Collizioni di prove per la lettura

**How to find information in the Prova collection**

There are 3 ways to find information in this collection:

- search for particular words
- search for *particular words* that appear in the text from the "search" page
- access publications by title
- access publications by filename

You can *search for particular words* that appear in the text from the "search" page. This is the first Page that comes up when you begin, and can be reached from other pages by pressing the search button.

You can *access publications by title* by pressing the *tiles* or *x-z* button. This brings up a list of books in alphabetic order.

You can *access publications by filename* by pressing the *filenames* button. This brings up a list of entries, sorted by original filename.

Internet

**Titoli in  
ordine  
alfabetico**

## **Titoli in ordine alfabetico**

Esempi – Il sistema Greenstone

# Metadati estratti automaticamente

The screenshot shows the Greenstone Librarian Interface with the title "Metadati estratti automaticamente". The interface has a menu bar with File, Edit, Metadata Sets, Hunt, Mirror, Gather, Enrich, Design, Export, and Create. A toolbar below the menu includes icons for Hunt, Mirror, Gather, Enrich, Design, Export, and Create. The main window displays a collection named "html\_small" containing files like "billbo.html", "Frodo.html", "merry.html", "pippin.html", and "sam.html". A sidebar on the left lists various metadata elements such as dc.Title, dc.Creator, dc.Subject, dc.Date, dc.Type, dc.Format, dc.Identifier, dc.Source, dc.Language, dc.Coverage, dc.Scripts, ex.Encoding, ex.Language, ex.Title, and ex.HTMLPlug. A tooltip on the right explains that the selected element, ex.Encoding, belongs to the "ex.Encoding" metadata set and can be extracted automatically using the 'Create' tab.

78

# Metadati assegnati manualmente

The screenshot shows the Greenstone Librarian Interface with the title "Metadati assegnati manualmente". The interface has a menu bar with File, Edit, Metadata Sets, Hunt, Mirror, Gather, Enrich, Design, Export, and Create. A toolbar below the menu includes icons for Hunt, Mirror, Gather, Enrich, Design, Export, and Create. The main window displays a collection named "html\_small" containing files like "billbo.html", "Frodo.html", "merry.html", "pippin.html", and "sam.html". A sidebar on the left lists various metadata elements such as dc.Title, dc.Creator, dc.Subject, dc.Date, dc.Type, dc.Format, dc.Identifier, dc.Source, dc.Language, dc.Coverage, dc.Rights, ex.Encoding, ex.Language, ex.Title, and ex.HTMLPlug. A tooltip on the right explains that the selected element, "Il signore degli anelli", belongs to the "Greenstone example library" metadata set and can be extracted manually using the 'Create' tab.

77

## Il modulo per la creazione dei metadati

The screenshot shows the Greenstone Librarian Interface with the title "Il modulo per la creazione dei metadati". The interface has a menu bar with File, Edit, Metadata Sets, Hunt, Mirror, Gather, Enrich, Design, Export, and Create. A toolbar below the menu includes icons for Hunt, Mirror, Gather, Enrich, Design, Export, and Create. The main window displays a collection named "html\_small" containing files like "billbo.html", "Frodo.html", "merry.html", "pippin.html", and "sam.html". A sidebar on the left lists various metadata elements such as dc.Title, dc.Creator, dc.Subject, dc.Date, dc.Type, dc.Format, dc.Identifier, dc.Source, dc.Language, dc.Coverage, dc.Rights, ex.Encoding, ex.Language, ex.Title, and ex.HTMLPlug. A tooltip on the right explains that the selected element, "Il signore degli anelli", belongs to the "Greenstone example library" metadata set and can be extracted manually using the 'Create' tab.

79