



Informatica **U**manistica

Lezione 10

Introduzione al sistema Greenstone

Pasquale Savino

ISTI - CNR



UNIVERSITÀ DI PISA

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

Biblioteche Digitali che usano Greenstone

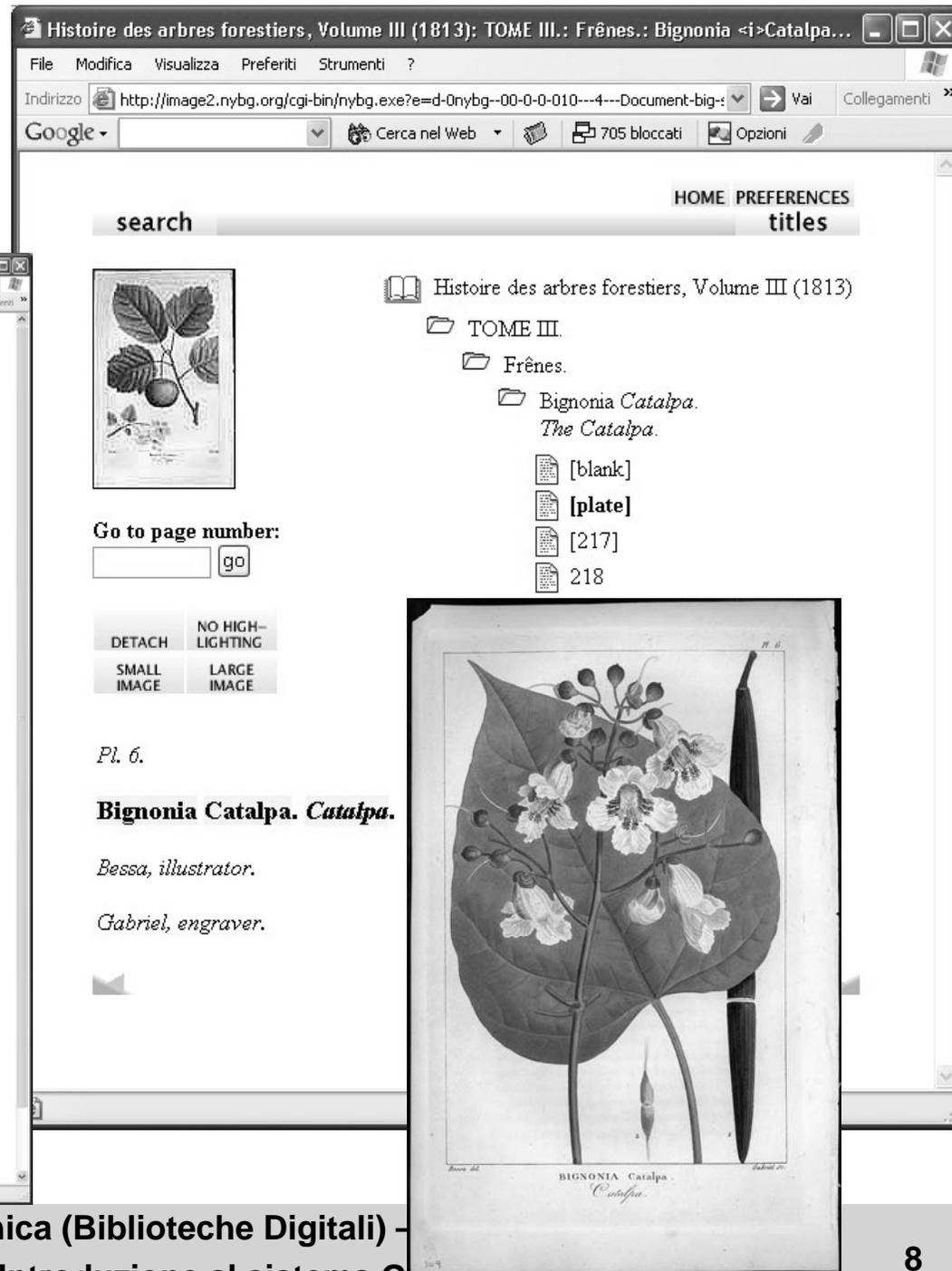
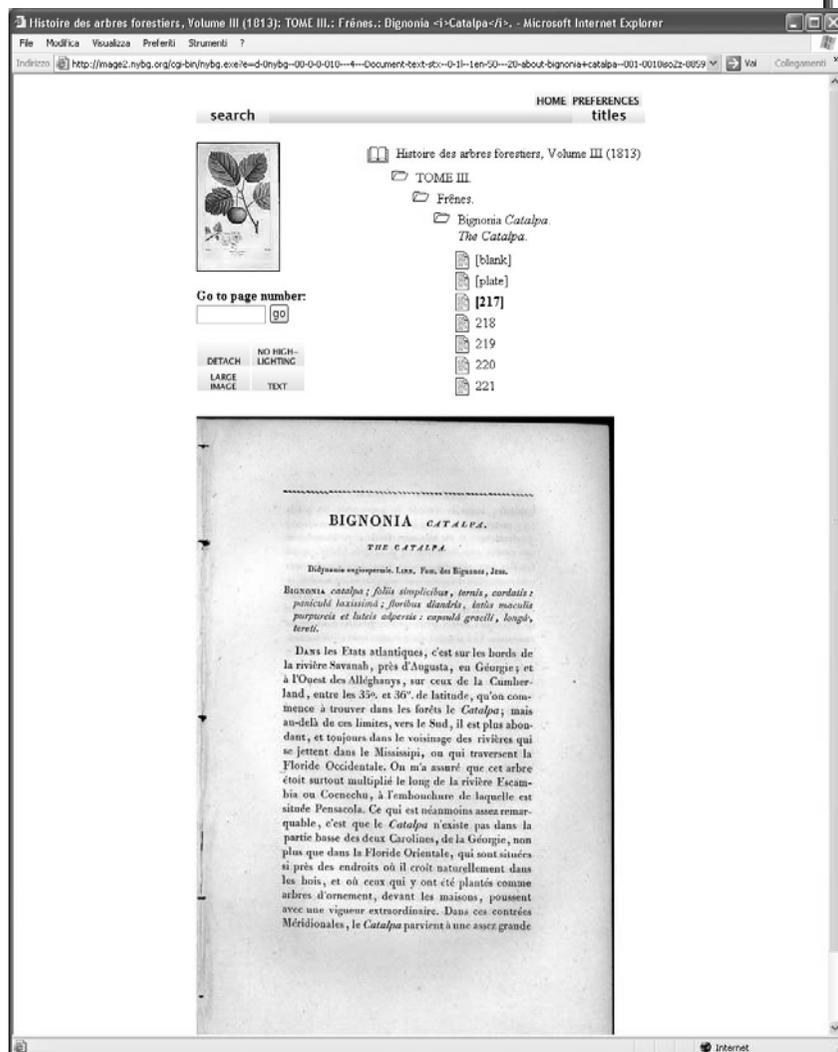
◆ Alcuni esempi

- 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/bsci/libr/rolen/page1.html>
- Project Gutenberg
→ <http://public.ibiblio.org/gsdll/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.aladin.wrlc.org/dl/>

◆ The New Zealand Digital Library

- <http://www.sadl.uleth.ca/nz/cgi-bin/library>

◆ New York Botanical Garden



◆ University of Chicago Library

■ Chopin Early Editions



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

The University of Chicago
LIBRARY

Chopin Early Editions  [HELP](#) [PREFERENCES](#)

[search](#) [titles](#) [uniform titles](#) [genres](#) [dedicatees](#)

Adieu à Varsovie : rondeau /

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

UniformTitle: Rondos, piano, op. 1, C minor
Title: Adieu à Varsovie : rondeau /
Title: "Adieu à Varsovie," rondeau in C min. op. 1
Composer: Chopin, Frédéric, 1810-1849.
Composer: Fontana, Julian, 1810-1865.
Publication Info.: London (No. 67, Frith Street, Corner of Soho Square) : Wessel & Co., importers and publishers of foreign music, [ca. 1840] Nouv. éd. corr.
Description: 13 p. of music ; 34 cm.
Notes: composé par Frederic Chopin.
 Caption title.
 "Edited by I. [i.e. J.] Fontana"--List t.p.
 Engraved. Price: 4/.
 Also available on the Internet.
Subjects: Rondos (Piano)
 Piano music.
Plate No.: W & Co. No. 1423
Plate No.: 1423
URI: <http://pi.lib.uchicago.edu/1001/dig/chopin/124>
Call No.: M25.C54R21 c.1

Page 2

Page 2

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

Testo strutturato gerarchicamente

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

Modello documenti gerarchico

Documento

Sezione

Titolo

I metadati possono essere associati ad ogni livello del documento

The screenshot shows a Microsoft Internet Explorer browser window displaying a website. The browser's address bar shows the URL: <http://www.sadl.uleth.ca/nz/cgi-bin/library?e=d-000-00---0frl2%2e2--00-0-0-0prompt-10---4-----0-11--1-en-50---20-about>. The website has a navigation menu with links for 'search', 'titles a-z', 'subjects', and 'organisations'. A search bar is visible. The main content area displays a book cover for 'Food First Curriculum' by Laurie Rubin. To the right of the cover is a table of contents with the following items:

- (introduction...)
- Acknowledgements
- To Harry Chapin
- Flowers are red
- Foreword
- Introduction
- Unit I - Why do people around the world do things in so many different ways?
- Unit II - Where does our food come from?
- Unit III - How do we get our food?
- Unit IV - Why are people hungry?
- Unit V - Who is hungry in the U.S.A.?
- Unit VI - What can we do?
- Resource guide
- About the institute for food and development policy...
- About the author...

Below the table of contents are buttons for 'EXPAND TEXT', 'EXPAND CONTENTS', and 'NO HIGH-LIGHTING'. The browser's status bar at the bottom shows 'Internet'.

Modello documenti gerarchico

Sezione

Sottosezione

Food First Curriculum for Grade 6 Students (FF, 1984, 146 p.): Unit III - How do we get our foo - Micr...

File Modifica Visualizza Preferiti Strumenti ?

Indietro Avanti Termina Aggiorna Pagina iniziale Cerca Preferiti Multimedia

Indirizzo <http://www.sadl.uleth.ca/nz/cgi-bin/library?e=d-000-00---0fnl2%2e2--00-0-0-0prompt-10---4-----0-1l--1-en-50---20-about> Vai

Google Cerca nel Web 703 bloccati Opzioni

HOME HELP PREFERENCES ORGANISATIONS

search titles a-z subjects

Food First Curriculum
by Laurie Rubin

Unit III - How do we get our food?

- Introduction
- Activity one - Visit to a supermarket
- Activity two - Processed foods-What's the difference?
- Activity three - What's in the package ?
- Activity four - Petroleum palace restaurant
- Activity five - Food ads-What's in them for me
- Activity six - What path did my food take?

EXPAND TEXT EXPAND CONTENTS
DETACH NO HIGH-LIGHTING

Operazione completata Internet

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/gsd/cgi-bin/library.cgi?a=p&p=about&c=gberg>

Testo non strutturato

The Abbot's Ghost, Or Maurice Treherne's Temptation _ A Christmas Story: 2 - Microsoft I...

File Modifica Visualizza Preferiti Strumenti ?

Indietro Avanti Termina Aggiorna Pagina iniziale Cerca Preferiti

Indirizzo <http://public.ibiblio.org/gsd/cgi-bin/library?e=d-000-00---0gberg--00-0-0-0prompt-10---> Vai Collegamenti

Google Cerca nel Web 710 bloccati Opzioni

HOME HELP PREFERENCES

search titles a-z authors a-z

The Abbot's Ghost, Or Maurice Treherne's Temptation _ A Christmas Story by Barnard, A. M.

page 2 (39 pages)

1 go to page 3

EXPAND TEXT DETACH

NO HIGH-LIGHTING

"It was General Snowdon and his charming wife. Maurice Treherne came while we were out, and I've not seen him yet, poor fellow!"

"Aye, you may well say that; his is a hard case, if what I heard is true. I'm not booked up in the matter, and I should be, lest I make some blunder here, so tell me how things stand, Major. We've a good half hour before dinner. Sir Jasper is never punctual."

"Yes, you've a right to know, if you are going to try your fortune with Octavia."

The major marched through the three drawing rooms to see that no inquisitive servant was eavesdropping, and, finding all deserted, he resumed his place, while young Annon lounged on a couch as he listened with intense interest to the major's story.

"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 Mère was very angry when she learned how the will was made."

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

Internet

Immagini di pagine di documenti

- ◆ **Immagini 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**

Immagini 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.sadl.uleth.ca/nz/cgi-bin/niupepalibrary?e=d-0niupepa--00-0-0-010--4----full---0-1l--1-it-50---20-about---00001-001-0utfZz-8-0&a=p&p=about>

N. 2: 3 - Microsoft Internet Explorer

File Modifica Visualizza Preferiti Strumenti ?

Indirizzo <http://www.sadl.uleth.ca/nz/cgi-bin/niupepalibrary?e=d-Oniuepa--00-0-0-010>

PAGINA PRINCIPALE

cerca serie

N. 2
15 Marzo 1861

pagina 3 (12 pagine)

2 vai a pagina 4

FULLSIZE IMAGE PREVIEW IMAGE NUOVA FINESTRA

NON EVIDENZIARE

the Rev Mr. Buller, the Rev Mr. Stack, the Resident Magistrate, Treasurer, Captain Scott J. Brittain, Esq, and other gentlemen in the of the Province, and in the civilisation and advancement subjects, occupied the platform. The Town Hall was filled with about forty five Maories present, of whom many were chiefs

The proceedings commenced by the Deputy Superintendent to the Maories which was interpreted Para, graph by paragraph by Rangiora, to the following effect

Welcome, my Maori friends I bid you welcome in the name of Canterbury. I am glad to see you here the more so because it that I have called this meeting, in order that you may have an expressing your loyalty to our Queen, and your friendly feeling

It is a grievous thing to see the Maori and the white man at wars had ceased in these islands and that there would have been the two races. But, alas it is not so.

It is not for me to enter upon the causes of the disputes which native tribes in the Northern Island to rebel against the sovereign course cannot but result in the ruin of those who have been so

Your presence here today shows that you have no sympathy have done well in coming here, to shake hands with the Pakeha that you have no bad feelings in your hearts towards them, and

N. 2: 3 - Microsoft Internet Explorer

File Modifica Visualizza Preferiti Strumenti ?

Indirizzo <http://www.sadl.uleth.ca/nz/cgi-bin/niupepalibrary?e=d-Oniuepa--00-0-0-010---4---text---0-11--1-it-50---20-about---00>

PAGINA PRINCIPALE AIUTO IMPOSTAZIONI NZDL

cerca serie date

N. 2
15 Marzo 1861

pagina 3 (12 pagine)

2 vai a pagina 4

TEXT PREVIEW IMAGE NUOVA FINESTRA

TE MANUHIRI TUARANGI AND MAORI INTELLIGENCER. 3

the Rev. Mr. Buller, the Rev. Mr. Stack, the Resident Magistrate, the Provincial Treasurer, Captain Scott, J. Brittain, Esq., and other gentlemen taking prominent interest in the progress of the Province, and in the civilization and advancement of their Maori fellow-subjects, occupied the platform. The Town Hall was filled with spectators. There were about forty-five Maories present, of whom many were chiefs.

The proceedings commenced by the Deputy Superintendent reading an address to the Maories, which was interpreted paragraph by paragraph by the Rev. J. W. Stack of Rangiora, to the following effect:—

Welcome, my Maori friends! I bid you welcome in the name of the Pakehas of Canterbury!

Pura (Minita), ko Te Taka (Minita), ko te Kai-whakawa, ko Te Mahimana, ko Kapene Kote, ko Te Pirirani, me etahi atu, ko nga tangata whakahaere tikanga o taua whenua, ko nga tangata e rapu whakaarona mo te Maori kia tupu haere ia i runga i te pai. Kapi tonu te whare i nga Pakeha me nga Maori, kei te whakarongo. Huihuia katoatia nga Maori, e wha tekau ma rima, he rangatira te nuinga.

Timataia ana te korero e Te Pouene, ko Te Taka te kai-whakamaori. Tenei ianei tana korero:—

Tena koutou, e aku hoa Maori. Ka aroha atu ahau, me matou katoa nga Pakeha o Katapere, ki a koutou.

Hari ana taku ngakau, mo koutou ka kitea mai, ina hoki, na koutou ake te whakaaro i karangatia ai e ahau tenei runanga. kia

Internet sistema Greenstone

Altri tipi di documenti

◆ Audio ed immagini fotografiche

- Esempio: Oral History Collection, Mirabilia Vicomercati
 - <http://www.sadl.uleth.ca/nz/cgi-bin/library?a=p&p=about&c=ohist>
 - <http://www.mirabiliavicomercati.org/gsd/cgi-bin/library?a=p&p=about&c=Foto&l=it>

◆ Video

- Esempio: Music Video Collection
 - <http://www.sadl.uleth.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.sadl.uleth.ca/nz/cgi-bin/music/musiclibrary>

◆ Lingue multiple

- Interfaccia e documenti
- Esempio: Arabic Collection
 - <http://www.sadl.uleth.ca/nz/cgi-bin/library?a=p&p=about&c=arabic&l=ar&nw=windows-1256>

Audio

The screenshot shows a web browser window with the title "Friday night was a big night in Hamilton" : the Golden Mile during the 1940s: Memories o...". The address bar shows the URL: http://www.sadl.uleth.ca/nz/cgi-bin/library?e=d-000-00---0ohist--00-0-0-0prompt-10---4-. The browser interface includes a menu bar (File, Modifica, Visualizza, Preferiti, Strumenti, ?) and a search bar. The page content features navigation links: search, titles a-z, HOME, HELP, PREFERENCES, and authors a-z. On the left, there are buttons for DETACH, NO HIGH-LIGHTING, and EXPAND TEXT. The main content area lists various topics with document icons: Background, Location, Homelife, Entertainment, Education, **Memories of WWII**, Education, Employment, Entertainment, Fashion, Awareness of local events, and Hamilton. Below this list, there is a section titled "Memories of WWII" with a speaker icon, containing the text: "Remembers Patriotic Hut in Garden Place, patriotic parades, end of the war, 'That was a day everyone went absolutely mad... all the teachers went mad and everybody sort of zigzagged or lacongad or something... down from the school, down Ward Street and right down into town, everybody got lost. I think it was one day we were allowed just to do as we liked'". The browser's status bar at the bottom shows "Internet" and a page number "20".

Immagini

Mirabilia Vicomercati - Archivi - Archivio Fotografico - Microsoft Internet Explorer

File Modifica Visualizza Preferiti Strumenti ?

Indirizzo <http://www.mirabiliavicomercati.org/gsd/cgi-bin/library?e=d-000-00---0Foto--00-0-0-0prompt-10> Vai Collegamenti >>

Mirabilia Vicomercati

Archivi

Archivio Fotografico PAGINA PRINCIPALE AIUTO IMPOSTAZIONI

titoli

cerca titoli autori soggetti persone date

A B C D F G I L M N O P R S T U V

 *Abitazioni della casa a corte di Via San Gerolamo*

 *Abitazioni della casa a corte di Via San Gerolamo*

 *Abitazioni 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*

 *Abside della Chiesa di Santo Stefano*

 *Abside della Chiesa di Santo Stefano*

Internet

Musica

Search page - Microsoft Internet Explorer

Indirizzo <http://www.sadl.uleth.ca/nz/cgi-bin/music/musiclibrary?e=p-dp-or-000-00---0folkfull--00-0-c>

HOME HELP PREFERENCES

folkfull search

search titles a-z browse

Symbolic Input Raw Audio Input

Add a rest

Duration

Information: Ready to record

Begin Search

Titles A-Z - Microsoft Internet Explorer

Indirizzo <http://www.sadl.uleth.ca/nz/cgi-bin/music/musiclibrary?e=d-dp-or-000-00---0folkfull--00-0-c>

HOME HELP PREFERENCES

folkfull titles a-z

search titles a-z browse

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z - Y 0-9

- A A A DER WINTER DER IST DA
- ABDUL ABULBUL AMEER
- A-Beggin' I Will Go
- ABEND SINKT STILL AUF DIE FLUR
- ABEND WIRD ES WIEDER
- Abendlied
- Abendlust
- Abendlust
- Abendreihen
- Abendreihen
- Abendreihen
- Abendreihen
- Abendreihen mit Kranzsingen
- Aber Montag, da ist a Knoeditag, - wenss nur all Tag Montag Knoeditag waere
- Abfertigung
- Abgewiesene Freier (Drei Gesellen) - Es war'n mal drei junge Gesellen

Audio Playback: Abendlust - Microsoft Internet Explorer

Applet meldexout started

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

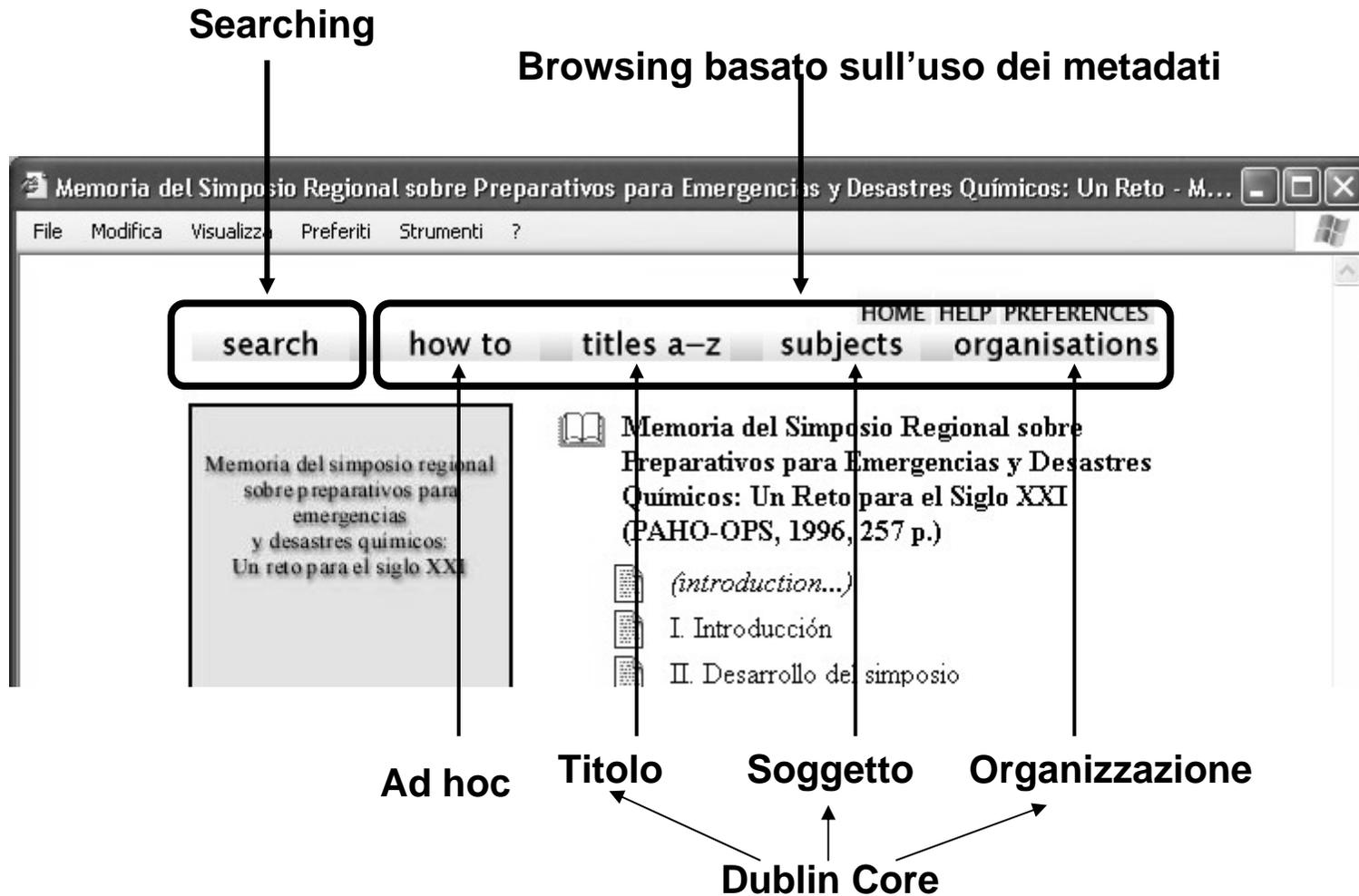
◆ MARC

◆ 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 e browsing



Searching

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

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

◆ 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 le stesse 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

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**

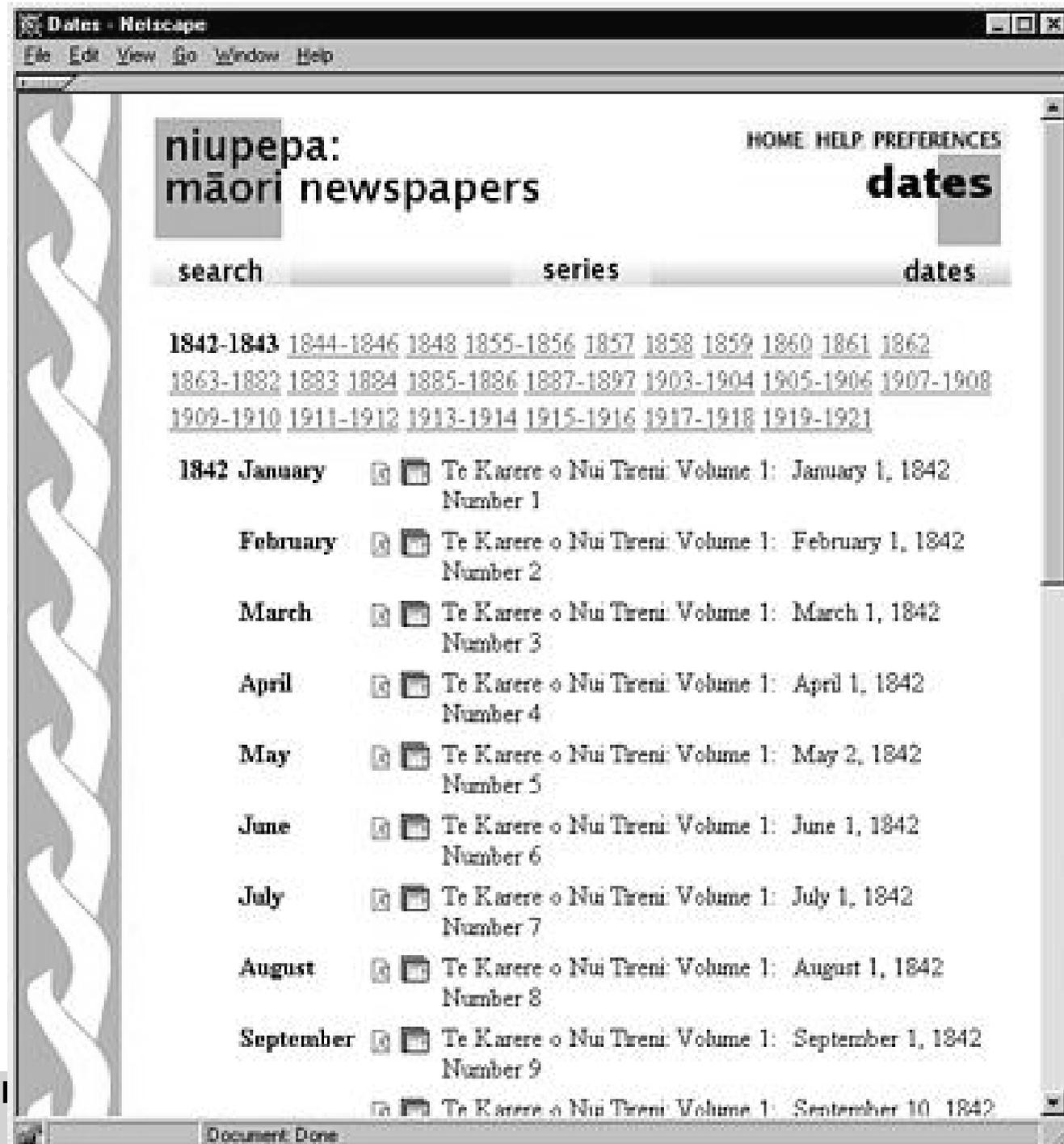
Browsing

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



Browsing

◆ Browsing per data



niupepa:
māori newspapers

HOME HELP PREFERENCES

dates

search series dates

[1841-1843](#) [1844-1846](#) [1848](#) [1855-1856](#) [1857](#) [1858](#) [1859](#) [1860](#) [1861](#) [1862](#)
[1863-1882](#) [1883](#) [1884](#) [1885-1886](#) [1887-1897](#) [1903-1904](#) [1905-1906](#) [1907-1908](#)
[1909-1910](#) [1911-1912](#) [1913-1914](#) [1915-1916](#) [1917-1918](#) [1919-1921](#)

1842 January   Te Karere o Nui Tirenī Volume 1: January 1, 1842
Number 1

February   Te Karere o Nui Tirenī Volume 1: February 1, 1842
Number 2

March   Te Karere o Nui Tirenī Volume 1: March 1, 1842
Number 3

April   Te Karere o Nui Tirenī Volume 1: April 1, 1842
Number 4

May   Te Karere o Nui Tirenī Volume 1: May 2, 1842
Number 5

June   Te Karere o Nui Tirenī Volume 1: June 1, 1842
Number 6

July   Te Karere o Nui Tirenī Volume 1: July 1, 1842
Number 7

August   Te Karere o Nui Tirenī Volume 1: August 1, 1842
Number 8

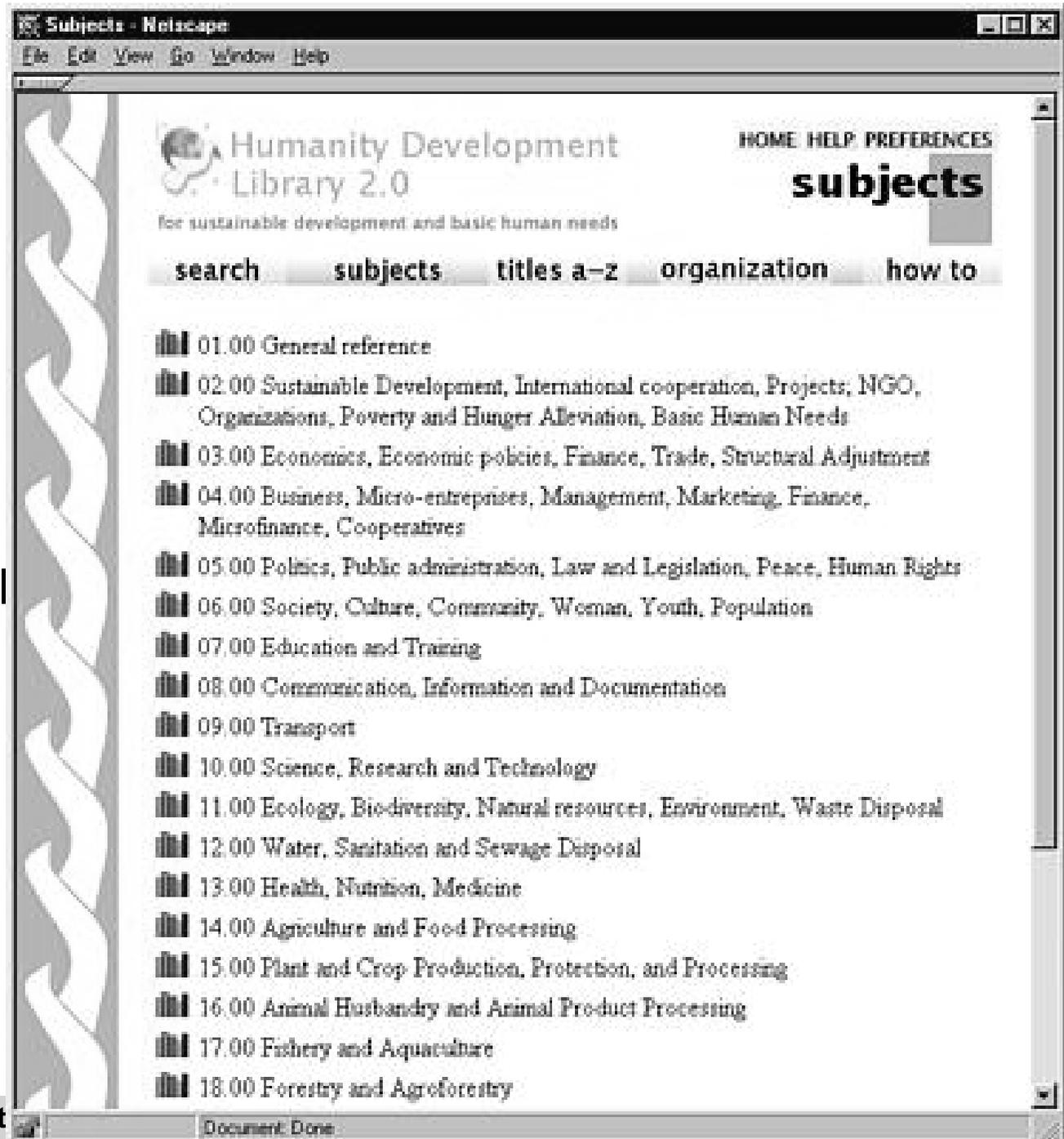
September   Te Karere o Nui Tirenī Volume 1: September 1, 1842
Number 9

  Te Karere o Nui Tirenī Volume 1: September 10, 1842

Document Done

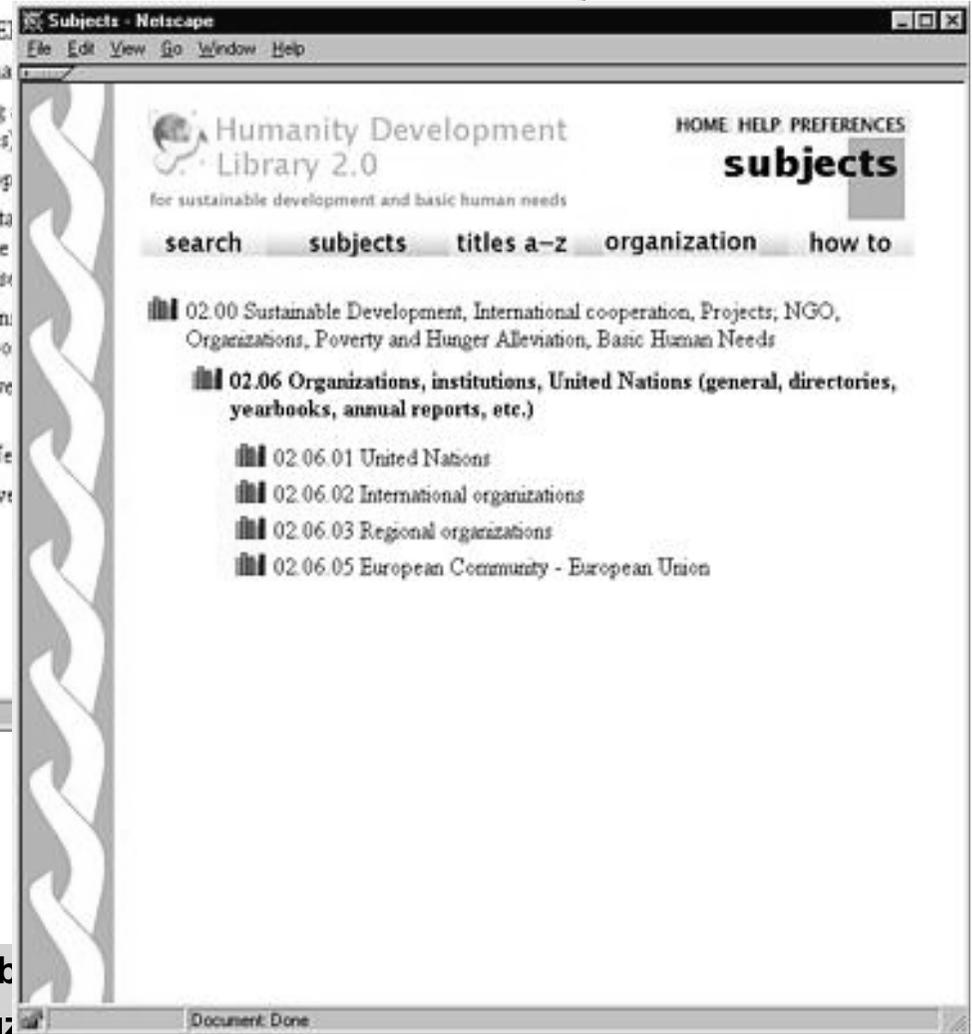
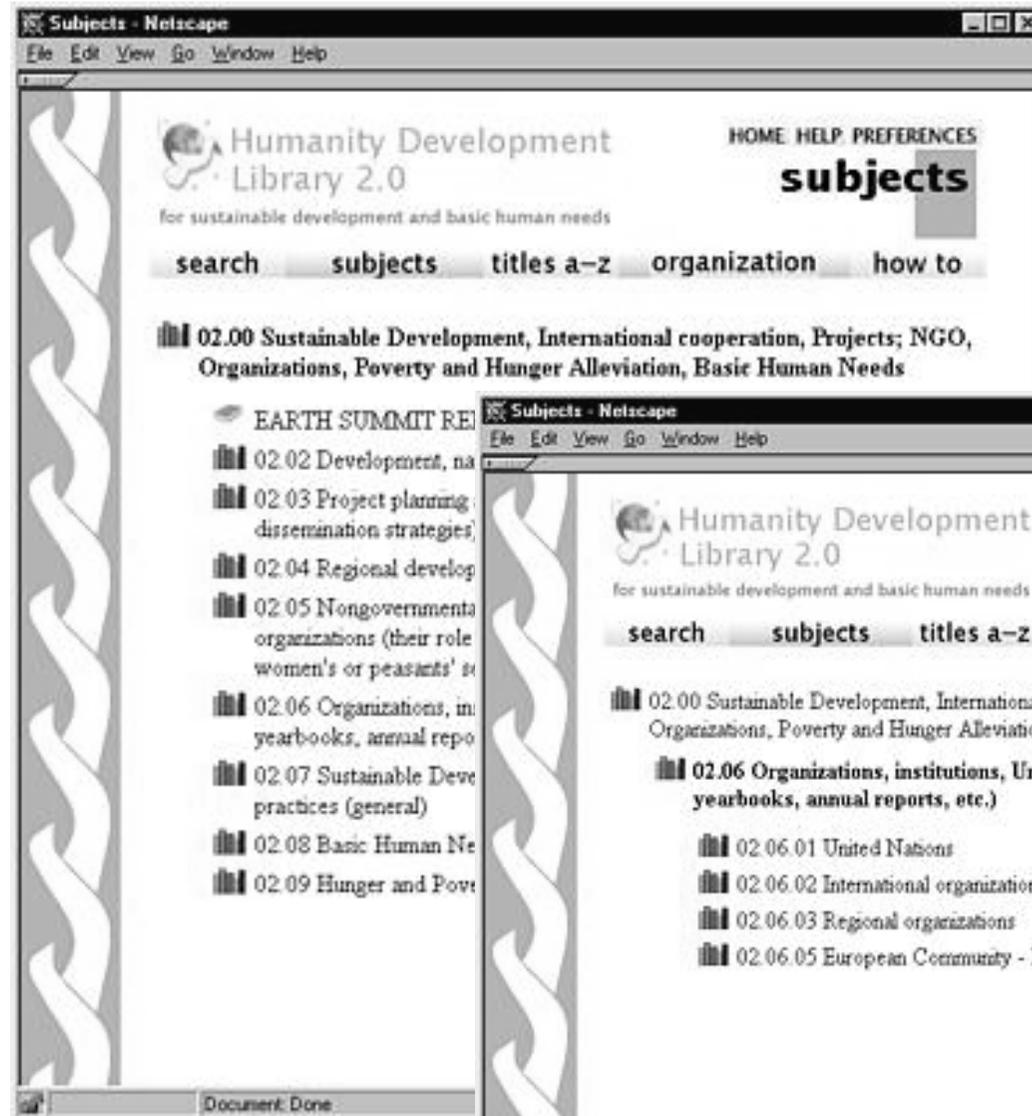
Browsing

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



Browsing

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



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.sadl.uleth.ca/nz/cgi-bin/library?a=p&p=about&c=fi1998>
 - <http://www.sadl.uleth.ca/nz/cgi-bin/library?a=p&p=about&c=allshake>

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

phind classifier demo

HOME HELP PREFERENCES

topics

search titles a-z language topics

Search for Previous Next

locust (first 10 of 102 phrases)

| | docs | freq |
|---------------------|------|------|
| Desert locust | 719 | 1767 |
| locust numbers | 127 | 289 |
| migratory locust | 89 | 192 |
| locust operations | 84 | 112 |
| locust infestations | 64 | 110 |
| locust situation | 66 | 94 |
| locust damage | 50 | 85 |
| locust control | 45 | 84 |
| locust infestation | 46 | 78 |
| forecast locust | 25 | 59 |

Get more phrases

Desert locust situation (first 10 of 102 phrases)

| | docs | freq |
|----------------------------|------|------|
| Desert locust situation | 140 | 188 |
| Desert locust adults | 99 | 153 |
| solitary Desert locust | 60 | 112 |
| Desert locust activity | 75 | 103 |
| Desert locust infestations | 60 | 88 |
| Desert locust control | 41 | 59 |
| Desert locust breeding | 39 | 49 |
| FAO Desert locust | 44 | 45 |
| Isolated Desert locust | 32 | 40 |
| Desert locust Bulletin | 32 | 38 |

Get more phrases

phind classifier demo

HOME HELP PREFERENCES

search titles a-z language topics

Search for Previous Next

locust (first 10 of 102 phrases)

| | docs | freq |
|---------------------|------|------|
| Desert locust | 719 | 1767 |
| locust numbers | 127 | 285 |
| migratory locust | 89 | 192 |
| locust operations | 84 | 112 |
| locust infestations | 84 | 110 |
| locust situation | 66 | 94 |
| locust damage | 50 | 85 |
| locust control | 45 | 84 |
| locust infestation | 46 | 76 |
| forecast locust | 25 | 59 |
| Get more phrases | | |

Desert locust (first 10 of 82 phrases, first 10 of 719 documents)

| | docs | freq |
|--|------|------|
| Desert locust control | 41 | 59 |
| Desert locust breeding | 39 | 49 |
| FAO Desert locust | 44 | 45 |
| Isolated Desert locust | 22 | 40 |
| Desert locust Bulletin | 32 | 38 |
| Get more phrases | | |
| Desert Locust Information Service of FAO: Locust FAQs | | 38 |
| JUST CONTROL SITUATION: MEASURES TAKEN AND FURTHER ACTION REQUIRED | | |
| Desert Locust Information Service of FAO: Locust books | | 20 |
| FAO - COMMITTEE ON AGRICULTURE | | 15 |
| PROGRAMME of WORK and BUDGET 30th Session of the Conference | | 13 |

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

Quindi posso
visualizzare il
documento

The screenshot shows a Netscape browser window with the following elements:

- Browser title: `_testTopicpage_ - Netscape`
- Menu bar: File Edit View Go Window Help
- Page navigation: HOME HELP PREFERENCES
- Search and navigation: search titles a-z language topics
- Logo: **LOCUST** (with a globe icon) and **DESERT LOCUST INFORMATION SERVICE** of the **MIGRATORY PESTS GROUP**
- Navigation menu: Locust Mapper News Archives Plagues Activities Books FAQ
- Section header: **Frequently Asked Questions (FAQs) about Desert Locusts**
- List of 25 frequently asked questions, each with a blue underlined link:

- [What is the difference between locusts and grasshoppers?](#)
- [What is a Desert Locust?](#)
- [What countries are affected by the Desert Locust?](#)
- [Do Desert Locust plagues occur with any regularity?](#)
- [How long does a Desert Locust live?](#)
- [How many eggs does a Desert Locust female produce?](#)
- [How far and how fast can Desert Locusts migrate?](#)
- [How big are swarms and how many locusts are there in a swarm?](#)
- [What percentage of the Desert Locust's exoskeleton is chitin?](#)
- [How much food can a Desert Locust eat?](#)
- [What is the relationship between locusts and ecology?](#)
- [Why do locusts change their behaviour?](#)
- [Are there other important species of locusts?](#)
- [Can locusts hurt humans?](#)
- [How can locusts be controlled?](#)
- [Who carries out locust control operations?](#)
- [Are there any non-chemical ways to kill locusts?](#)
- [Can locusts be detected by satellites?](#)
- [Why are Desert Locust so difficult to control?](#)
- [Do people really eat locusts?](#)
- [What is a Desert Locust composed of?](#)

At the bottom of the browser window, the status bar shows "Document Done".

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
- Funziona solo in locale

◆ Uso della Librarian Interface

- Funziona in locale
- Interfaccia per creare nuovi valori di metadati

Il “Collector” di Greenstone

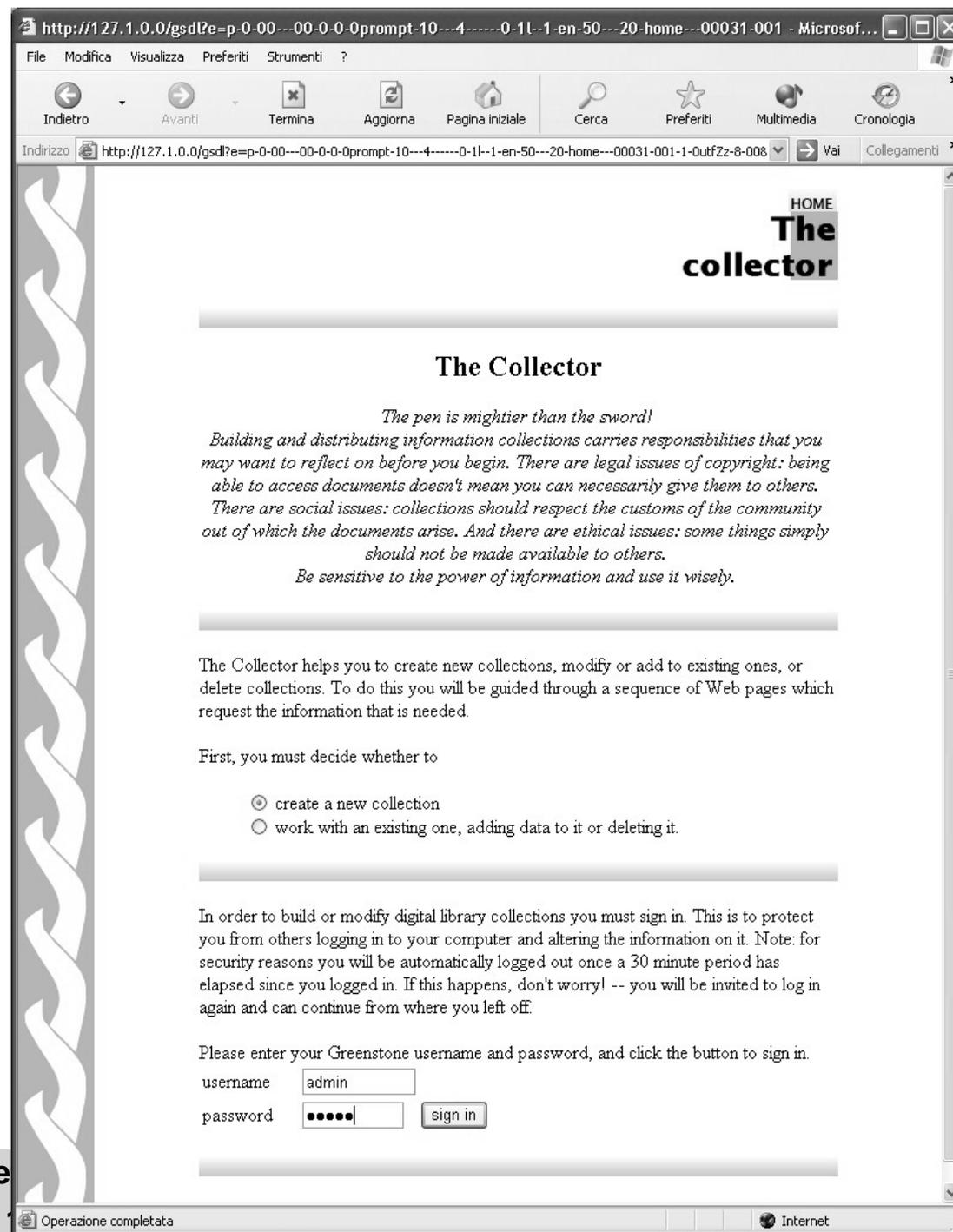
◆ Il Collector permette di

- 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



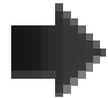
Creazione di una nuova collezione

- ◆ **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 informaizone associata 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

HOME
The collector

Collection information

When creating a new collection you need to enter some preliminary information about the source data. This process is structured as 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 collection title is a short phrase used throughout the digital library to identify the content of the collection. Example titles include "Computer Science Technical Reports" and "Humanity Development Library."

Contact email address:

This email address specifies the first point of contact for the collection. If the Greenstone software detects a problem, a diagnostic report is sent to this address. Enter an email address in its full form: name@domain.

About this collection:

This is a statement describing the principles governing what is included in the collection. It appears on the first page when the collection is presented.

Your position in the sequence is indicated by an arrow underneath--in this case, the "collection information" stage. To proceed, click the green "source data" button.

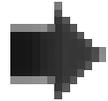
▶ collection information ▶ source data ▶ configure collection ▶ build collection ▶ view collection

▲

Operazione completata

Internet

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://

◆ **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

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

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

Dati da inserire nella collezione

HOME
The collector

Source data

You may base your collection on either

- The default structure
The new collection may contain html documents (.htm, .html), plain text documents (.txt, .text), MS Word documents (.doc), PDF documents (.pdf) or "m-box" format email documents (.mbx).
- An existing collection
The files in your new collection must be exactly the same type as those used to build the existing one.

Base the collection on

Input sources:

file://

http://

ftp://

MORE SOURCES

If you use file:// or ftp:// to specify a file, that file will be downloaded.

If you use http:// it depends on whether the URL gives you a normal web page in your browser, or a list of files. If a page, that page will be downloaded -- and so will all pages it links to, and all pages they link to, etc. -- provided they reside on the same site, below the URL.

If you use file:// or ftp:// to specify a folder or directory, or give a http:// URL that leads to a list of files, everything in the folder and all its sub-folders will be included in the collection.

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

Click one of the green buttons. If you are an advanced user you may want to adjust the collection configuration. Alternatively, go straight to the building stage. Remember, you can always revisit an earlier stage by clicking its yellow button.

Operazione completata Internet

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**

Modifica della configurazione

HOME
The collector

Configure collection

The building and presentation of your collection are controlled by specifications in a special "configuration file". Advanced users may want to alter the configuration settings.

If you are not an advanced user, just go to the bottom of the page.

To alter the configuration settings, edit the data that appears below. If you make a mistake, click on "Reset" to reinstate the original configuration settings.

```
creator      savino@isti.cnr.it
maintainer   savino@isti.cnr.it
public       true

indexes      document:text document:Title document:Source
defaultindex document:text

plugin       ZIPPlug
plugin       GAPlug
plugin       TEXTPlug
plugin       HTMLPlug
plugin       EMAILPlug
plugin       PDFPlug
plugin       RTFPlug
plugin       WordPlug
plugin       PSPlug
plugin       ArcPlug
plugin       RecPlug
```

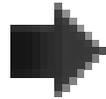
[Reset](#)

collection information ▶ source data ▶ **configure collection** ▶ build collection ▶ view collection

Operazione completata Internet

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 progredisce l'operazione**

Costruzione della collezione

http://127.1.0.0/gsd - Microsoft Internet Explorer

File Modifica Visualizza Preferiti Strumenti ?

Indirizzo http://127.1.0.0/gsd Vai Collegamenti >>

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.

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

Operazione completata

Costruzione della collezione

http://127.1.0.0/gsd - Microsoft Internet Explorer

File Modifica Visualizza Preferiti Strumenti ?

Indirizzo http://127.1.0.0/gsd Vai Collegamenti >>

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.

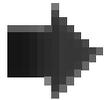
Collection built successfully.

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

Operazione completata Internet

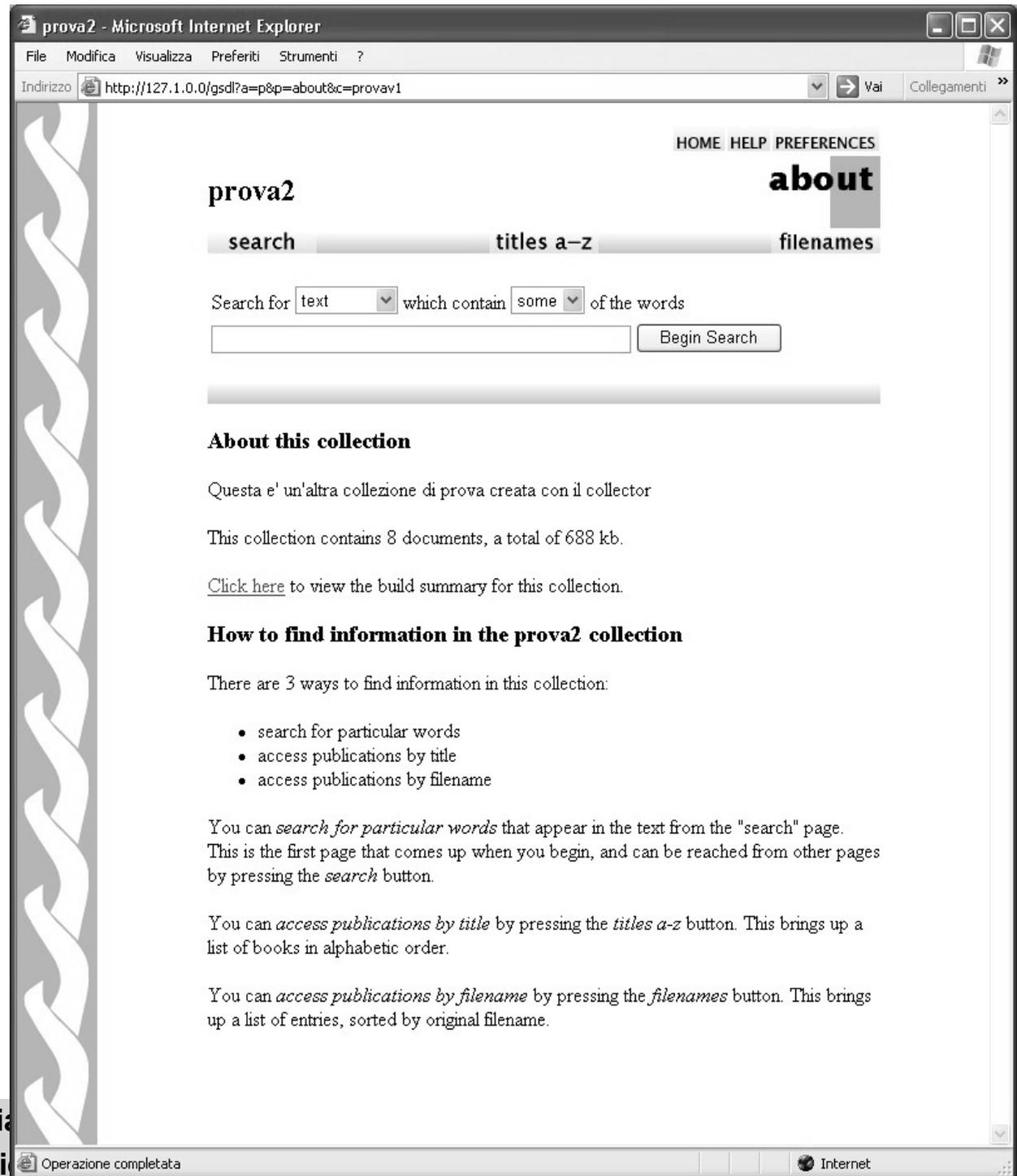
Passi per la costruzione della collezione

- 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



The screenshot shows a Microsoft Internet Explorer browser window displaying the 'prova2' web application. The browser's address bar shows the URL: `http://127.1.0.0/gsd?l=a&p=about&c=provav1`. The page features a navigation menu with 'HOME', 'HELP', and 'PREFERENCES'. The main content area is titled 'prova2' and includes three buttons: 'search', 'titles a-z', and 'filenames'. Below these buttons is a search form with a dropdown menu set to 'text', a second dropdown set to 'some', and a 'Begin Search' button. The page also contains sections for 'About this collection' and 'How to find information in the prova2 collection', providing instructions on how to use the search and navigation features.

prova2

HOME HELP PREFERENCES

search titles a-z filenames

Search for which contain of the words

About this collection

Questa e' un'altra collezione di prova creata con il collector

This collection contains 8 documents, a total of 688 kb.

[Click here](#) to view the build summary for this collection.

How to find information in the prova2 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 alphabetic order.

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

Operazione completata

Internet

Lista dei
documenti
inseriti nella
nuova
collezione

The screenshot shows a Microsoft Internet Explorer browser window titled "Titles A-Z - Microsoft Internet Explorer". The address bar contains the URL "http://127.1.0.0/gsd?e=p-0-00-provav1--00-0-0-0prompt-10---4-----0-1--1-en-50---20-about---00031-001-1-Outf". The browser interface includes a menu bar with "File", "Modifica", "Visualizza", "Preferiti", "Strumenti", and "?". The main content area displays a web application with a header "HOME HELP PREFERENCES" and a main title "titles a-z". Below the title are three tabs: "search", "titles a-z", and "filenames". The "titles a-z" tab is active, showing a list of documents. Each document entry includes a file icon, a title, and a filename in parentheses. The list contains the following items:

- 1992-96 Working Papers
(word04.doc)
- 1997-00 Listing of Working Papers
(word01.doc)
- biblio_for_dl_scientometrics.do
(pdf03.pdf)
- Distributing Digital Libraries on the Web, CD-ROMs, and Intranets: Same information, same...
(pdf02.pdf)
- Greenstone: A Comprehensive Open-Source
(word03.doc)
- Greenstone: A Comprehensive Open-Source Digital Library Software System
Ian H. Witten, *Rodger J...
(pdf01.pdf)
- Information Seeking, Retrieving, Reading, and Storing Behaviour of Library-Users.
(word02.doc)
- Scalable browsing for large collections: a case study Gordon W. Paynter, 1 Ian H. Witten, 1 Sally...
(pdf04.pdf)

The browser's status bar at the bottom shows "Operazione completata" and "Internet".

Esercitazione – Parte I

Creare una collezione usando il Collector