

## Sommario

- ◆ **Processo di funzionamento di importazione e building di una collezione**
  - Import
  - Build
- ◆ **Il configuration file**
  - ◆ **Uso di**
    - 1. Plug-in
    - 2. Classifiers
    - 3. Indici
- ◆ **Formattazione delle pagine web**

Pasquale Savino – ISTI-CNR	Biblioteche Digitali Esempi – Il sistema Greenstone	1
	Biblioteche Digitali Esempi – Il sistema Greenstone	2

## Greenstone

## Funzionalità avanzate

- ◆ **Vi è una distinzione tra**
  - COSTRUIRE una collezione
  - FORNIRE informazioni agli utenti
- ◆ **È la stessa distinzione che esiste tra il ‘compile-time’ ed il ‘runtime’ nei linguaggi di programmazione**
- ◆ **La fase di costruzione è necessaria per preparare tutte le strutture dati che vengono poi utilizzate nella fase di delivery delle informazioni**

Pasquale Savino – ISTI-CNR	Biblioteche Digitali Esempi – Il sistema Greenstone	3
	Biblioteche Digitali Esempi – Il sistema Greenstone	4

## Digital Library Collections

## Costruzione di una collezione

- ◆ Il processo che consiste nel prendere un insieme di documenti ed i metadati che li descrivono e creare tutti gli indici e le strutture dati che ne supportano la ricerca (search), il browsing, e la visualizzazione

## Costruzione di una collezione

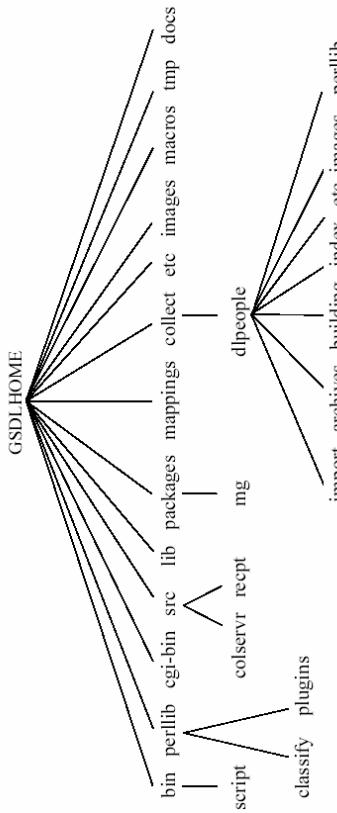
- ◆ La costruzione di una collezione prevede quattro fasi
  - Make
    - Creare uno scheletro di strutture e di file nel quale verranno inseriti i dati della collezione
  - Import
    - Importare i documenti ed i metadati e convertirli nel formato Greenstone
  - Build
    - Costruire gli indici e le strutture dati richieste
  - Install
    - Rendere operativa la collezione

Biblioteche Digitali  
Esempi – II sistema Greenstone 5  
Pasquale Savino – ISTI-CNR

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

## Make

- ◆ Vengono create le seguenti directory



Biblioteche Digitali  
Esempi – II sistema Greenstone 7  
Pasquale Savino – ISTI-CNR

Biblioteche Digitali  
Esempi – II sistema Greenstone 8  
Pasquale Savino – ISTI-CNR

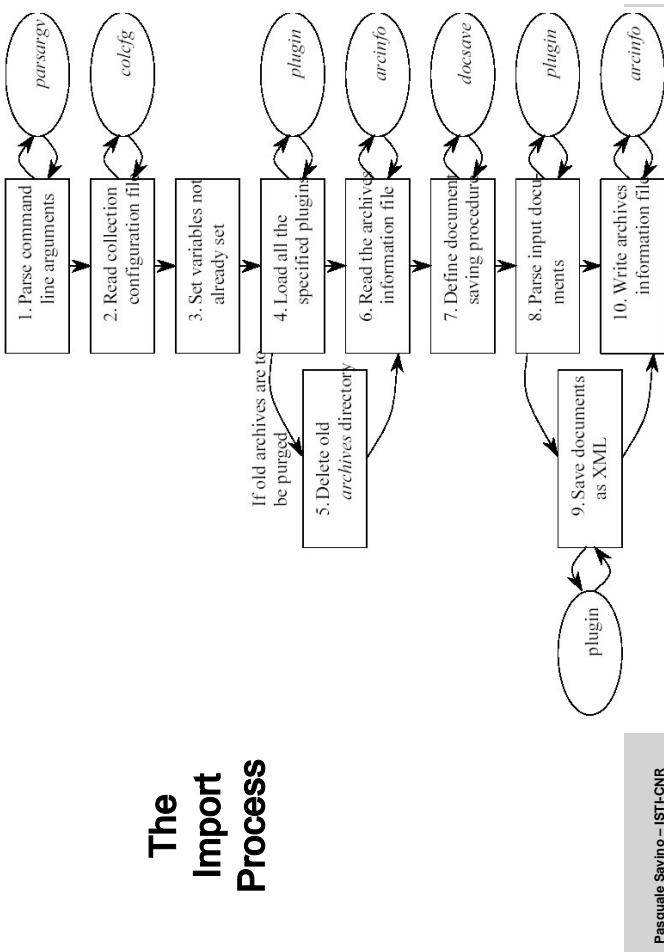
*bin* Executable code, including binaries in the directory with your O/S name.  
*bin/script* Perl scripts used for creating and building collections (for example *import.pl* and *buildcoll.pl*). To get a description of any of these programs, type their name at the command prompt.  
*perl/lib* Perl modules used at import and build time (plugins, for example).  
*perl/lib/plugins* Perl code for document processing plugins.  
*perl/lib/classify* Perl code for classifiers (for example the AZList code that makes a document list based on the alphabetical order of some attribute).  
*cgi-bin* All Greenstone CGI scripts, which are moved to the system cgi-bin directory.  
*tmp* Directory used by Greenstone for storing temporary files.  
*etc* Configuration files, initialisation and error logs, user authentication databases.  
*src* C++ code used for serving collections via a web server.  
*src/collserv* C++ code for serving collections answering queries and the like.  
*src/rept* C++ code for getting queries from the user interface and formating query responses for the interface.  
*packages* Source code for non-Greenstone software packages that are used by Greenstone.  
*packaging* The source code for MG, the compression and indexing software used by Greenstone.  
*mappings* Unicode translation tables (for example for the GB Chinese character set).  
*macros* The macro files used for the user interface.  
*collect* Collections being served from this copy of Greenstone  
*lib* C++ source code used by both the collection server and the recipient.  
*images* Images used in the user interface.  
*docs* Documentation.

# Import Process

- ◆ Brings documents and metadata into the system in a **standardized XML form**
- ◆ Original material placed in **import directory**
- ◆ Import process transforms it to files in the **archives directory**
- ◆ The original material can be **deleted**
  - Collection can be rebuilt from archive files
- ◆ New material added to collection by placing it in **import directory** and re-executing the import process
  - The new material finds its way into archives along with existing files
- ◆ To keep the source form of collections
  - Do not delete the archives
    - "Source" form can be augmented and rebuilt later

Pasquale Savino – ISTI-CNR  
Biblioteche Digitali  
Esempi – Il sistema Greenstone 9

## The Import Process



## Options for Import

<code>-verbosity</code>	Number 0-3	Control how much information about the process is printed to standard error: 0 gives a little, 3 gives lots.
<code>-archivedir</code>	Directory name	Specify where the Greenstone archive files are stored—that is, where <code>import.pl</code> puts them and where <code>buildcol.pl</code> finds them. Defaults to <code>GSDLHOME/collect/col_name/archives</code>
<code>-maxdocs</code>	Number >0	Indicates the maximum number of documents to be imported or built. Useful when testing a new collection configuration file, or new plugins.
<code>-collectdir</code>	Directory name	Specify where the collection can be found. Defaults to <code>GSDLHOME/collect</code>
<code>-out</code>	Filename	Specify a file to which to write all output messages, which defaults to standard error (the screen). Useful when working with debugging statements.
<code>-keepold</code>	None	Do not remove the result of the previous import or build operation. In the case of import, do not remove the contents of the <code>archives</code> directory; when building, do not remove the content of the <code>building</code> directory.
<code>-debug</code>	None	Print plugin output to standard output.

## Additional Options for Import

<code>-importdir</code>	Directory name	Where material to be imported is found. Defaults to <code>GSDLHOME/collect/col_name/import</code> .
<code>-removeold</code>	None	Remove the contents of the <code>archives</code> directory before importing.
<code>-gzip</code>	None	Zip up the Greenstone archive documents produced by <code>import</code> ( <code>ZIP</code> plug must be included in the plug-in list, and <code>gzip</code> must be installed on your machine).
<code>-groupsize</code>	Number >0	Number of documents to group together into one Greenstone archive file, defaults 1 (that is, one document per file).
<code>-sortmeta</code>	Metadata tag name	Sort the documents alphabetically by the named metadata tag. However, if the collection has more than one group in the collection (i.e. <code>groupsize &gt; 1</code> ), this functionality is disabled.
<code>-OIDtype</code>	<code>hash</code> or <code>incremental</code>	Method of creating OIDs for documents: <code>hash</code> hashes the content but is slow; <code>incremental</code> simply assigns document numbers sequentially, and is faster.

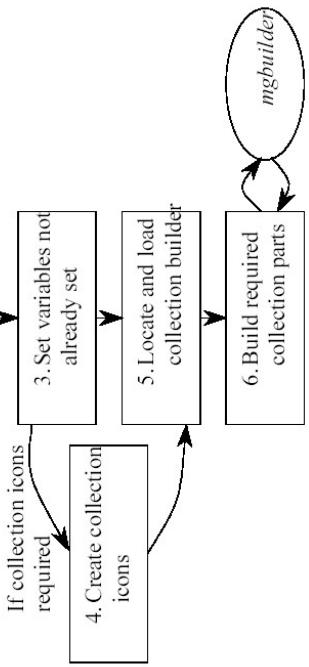
## The Build Process

- ◆ Creates the indexes and data structures that make the collection operational

- ◆ Indexes for the whole collection are build all at once

- Build process does not work incrementally
  - Adding new material to *archives* requires that entire collection be rebuilt (by issuing *buildcol.pl*)
    - Most collections can be rebuilt overnight

## The Build Process



## Options for Build

<i>-verbosity</i>	Number 0–3	Control how much information about the process is printed to standard error: 0 gives a little, 3 gives lots.
<i>-archivedir</i>	Directory name	Specify where the Greenstone archive files are stored—that is, where <i>import.pl</i> puts them and where <i>buildcol.pl</i> finds them. Defaults to <i>GSDLHOME/collect/col_name/archives</i>
<i>-maxdocs</i>	Number >0	Indicates the maximum number of documents to be imported or built. Useful when testing a new collection configuration file, or new plugins.
<i>-collectdir</i>	Directory name	Specify where the collection can be found. Defaults to <i>GSDLHOME/collect</i>
<i>-out</i>	Filename	Specify a file to which to write all output messages, which defaults to standard error (the screen). Useful when working with debugging statements.
<i>-keepold</i>	None	Do not remove the result of the previous import or build operation. In the case of import, do not remove the contents of the <i>archives</i> directory; when building, do not remove the content of the <i>building</i> directory.
<i>-debug</i>	None	Print plugin output to standard output.

## Additional Options for Build

<i>-builddir</i>	Directory name	Specify where the result of building is to be stored (defaults to <i>GSDLHOME/collect/col_name/building</i> ).
<i>-index</i>	Index name (e.g., <i>section&gt;Title</i> )	Specify which indexes to build. This defaults to all the indexes indicated in the collection configuration file.
<i>-allclassifications</i>	None	Prevent the build process from removing classifications that include no documents (for example, the "X" classification in titles if there are no documents whose titles start with the letter X).
<i>-create_images</i>	None	Create collection icons automatically (to use this, GIMP, and the Gimp Perl module, must be installed).
<i>-mode</i>	<i>all</i> , <i>compress_text</i> , <i>infodb</i> , or <i>build_index</i>	Determine what the build process is to do (defaults to <i>all</i> ). <i>All</i> does a full build, <i>compress_text</i> only compresses the document text, <i>infodb</i> creates a database of information pertaining to the collection—name, files, associated files, classification information and the like—and <i>build_index</i> builds the indexes specified in the collection configuration file or on the command line.
<i>-no_text</i>		Don't store compressed text. This option is useful for minimizing the size of the built indexes if you intend always to display the original documents at run-time.

Collection Configuration File

- ◆ **Collection Configuration File**
    - Defines the structure of a collection
    - Governs how the collection is built
    - Specifies how the collection will appear to users
  - ◆ **Ogni linea del Collection Configuration File è una coppia “attributo”, “valore”**

Collection Configuration File

Pasquale Savino - ISTI-CNR  
Biblioteche Digitali  
Esempi - Il sistema Greenstone  
17

Pasquale Savino - ISTI-CNR  
Biblioteche Digitali  
Esempi – Il sistema Greenstone  
18

Collection Configuration File [1/4]

<i>creator</i>	E-mail address of the collection's creator
<i>maintainer</i>	E-mail address of the collection's maintainer
<i>public</i>	Whether collection is to be made public or not
<i>beta</i>	Whether collection is beta version or not
<i>indexes</i>	List of indexes to build
<i>defaultIndex</i>	The default index
<i>subcollection</i>	Define a subcollection based on metadata
<i>indexSubcollections</i>	Specify which subcollections to index
<i>defaultSubcollection</i>	The default index subcollection
<i>languages</i>	List of languages to build indexes in
<i>defaultLanguage</i>	Default index language
<i>collectionName</i>	Defines collection-level metadata
<i>plugin</i>	Specify a plugin to use at build time
<i>format</i>	A format string (explained below)
<i>classify</i>	Specify a classifier to use at build time

Collection configuration file [2/4]

creator	<code>username@email.it</code>	Plugin da usare per convertire documenti nel formato Greenstone
maintainer	<code>username@email.com</code>	
public	true	
beta	false	Classificatore per creare una lista alfabetica di titoli

Plugin GAPPlug  
Plugin HTMlPlug -description\_tags -cover\_image  
Plugin WordPlug -description\_tags  
Plugin ArcPlug  
Plugin RecPlug -show\_progress -use\_metadata\_files

classify AZList metadata title

Pasquale Savino - ISTI-CNR      Biblioteche Digitali  
Esempi - II sistema Greenstone      19

Pasquale Savino - ISTI-CNR      Biblioteche Digitali      Esempi - Il sistema Greenstone  
20

Collection configuration file [3/4]

The diagram illustrates the relationship between two entities: **Collection** and **Document**.

- Collection** contains **CollectionMeta** and **CollectionContent**.
- CollectionMeta** includes **CollectionName**, **CollectionDescription**, and **CollectionIconCollection**.
- CollectionContent** includes **CollectionText**, **CollectionImage**, and **CollectionButtons**.
- CollectionButtons** includes **ExpandText**, **Expand**, **Contents**, **Detach**, and **Highlight**.

A callout box labeled "Formato di presentazione dei metadati" points to the **CollectionMeta** section.

A callout box labeled "Metadati della collezione" points to the **CollectionContent** section.

Subcollections [1/2]

- ◆ Greenstone permette di costruire sotto-collezioni, e di costruire indici per ognuna di esse.
  - ◆ Consideriamo una collezione costituita documenti testuali, alcuni tratti dal “Journal of Digital Libraries” ed altri no
  - ◆ Vogliamo creare due sotto-collezioni ed indici al livello di section

indexes section:text  
subcollection dl "title/^Journal of Digital Libraries/i"  
subcollection other "title/^Journal of Digital  
Libraries/i"  
indexsubcollections dl other dl,other

Collection configuration file [4/4]

The diagram illustrates the relationship between two entities: **Collectionmeta** and **Collectionextra**.

- Collectionmeta** contains the attribute `collectiondescription` (highlighted in red).
- Collectionmeta** contains the attribute `collectionname` (highlighted in blue).
- Collectionextra** contains the attribute `collectiondescription` (highlighted in red).
- Collectionextra** contains the attribute `collectionname` (highlighted in blue).

A curved arrow points from the `collectiondescription` attribute in **Collectionmeta** to the `collectiondescription` attribute in **Collectionextra**, indicating that the description of the collection is shared or derived.

Pasquale Savino - ISTI-CNR      Biblioteche Digitali  
Esempi - Il sistema Greenstone      22

Subcollections [2/2]

- ◆ Lo stesso meccanismo può essere utilizzato per creare indici per collezioni che contengono documenti in diverse lingue
  - ◆ La lingua del documento è un metadato (en per l'inglese, it per italiano, ecc.)

indexes      section:text    section:title    document:text  
Languages    it    en    fr

◆ Vengono creati indici separati per section text, section title, e document text per le tre diverse lingue (9 indici in totale)

Pastquale Savino - ISTI-CNR      Biblioteche Digitali  
Esempi - Il sistema Greenstone      23

Biblioteche Digitali  
Esempi – Il sistema Greenstone

Cross-collection searching

- ◆ In Greenstone è possibile effettuare ricerche su più collezioni, come se fossero costituite da una sola collezione
  - ◆ Questa funzionalità viene abilitata inserendo nel Collection Configuration File
    - supercollection col\_1 col\_2 ...
  - ◆ Nel caso che le collezioni siano denominate col\_1, col\_2, ecc.
  - ◆ Questa indicazione deve essere presente nel file di configurazione di tutte le collezioni coinvolte.

## Plug-ins

supercollection col\_1 col\_2 ...

**Il caso che le collezioni siano collection col\_1 col\_2 ...**

◆ Questa indicazione deve essere presente nel file di configurazione di tutte le collezioni coinvolte.

## Plug-ins

- ◆ I **plug-in sono moduli software che gestiscono**
    - Conversioni di formato
    - Estrazione di metadati
  - ◆ I **plug-in permettono di estendere le funzionalità di Greenstone**
    - È possibile sviluppare nuovi plug-in per estendere i tipi di documenti gestiti o i metadati che possono essere estratti
  - ◆ I **plug-in sono scritti nel linguaggio Perl. Sono tutti derivati da un plug-in base: BasPlug.**
  - ◆ **BasPlug** crea un nuovo documento archivio di Greenstone ed assegna un identificatore al documento
  - ◆ Maggiori informazioni su ogni plug-in si possono avere digitando “perl –S plugin.info.pl nome-plugin” alla linea comandi di windows

Plug-ins

- **Plug-ins do most of the work of the import process**
  - **Operate in the order in which they are listed in the *collect.cfg* file**
    - Input file is passed to each plug-in until one is found that can process it
    - **If there is no plug-in that can process a file, a warning is printed**
  - **Plug-ins determine the traversal of the subdirectory structure in the import directory**
    - *RecPlug* - processes directories, recurses through directory structures and passes the name through the plug-in list
    - *GAPPlug* – processes Greenstone Archive Format documents (in the archives directory structure)
    - *ArcPlug* – used during building, processes list of document OIDs produced during import (list is stored in *archives.infile*)

Biblioteche Digitali Esempi – Il sistema Greenstone	Pasquale Savino – ISTI-CNR	25
Biblioteche Digitali Esempi – Il sistema Greenstone	Pasquale Savino – ISTI-CNR	26

Biblioteche Digitali  
Economia, sistemi, contesti 27

**Biblioteche Digitali** 28  
Economia - Il Giacomo Gramsci

Pasquale Savino - ISTI-CNR	27	Pasquale Savino - ISTI-CNR	28
Biblioteche Digitali Fondazione Istituto Italiano della Cultura		Biblioteche Digitali Fondazione Istituto Italiano della Cultura	

## Plug-ins & Document Formats

- ◆ Plug-ins are specified in the collection configuration file
  - ◆ File name determines document format
    - ◆ Widely used document formats:

TEXTPlug	PSPlug
HTMLPlug	EMAILPlug
WORDPlug	BibTeXPlug
PDFPlug	ReferPlug
SRCPPlug	ImagePlug
ZIPPlug	

## Text Files

- ◆ TEXTPlug Plug-in
    - \*.txt
    - \*.text
  - ◆ Plain text file
  - ◆ Title metadata based on the first line of the file

Pasquale Savino – ISTI-CNR      Esempi – Il sistema Greenstone  
Biblioteche Digitali      29

Pasquale Savino – ISTI-CNR      Esempi – Il sistema Greenstone  
Biblioteche Digitali      30

HTML Files

- ◆ **HTMLPlug Plug-In**
    - Imports HTML files
    - Title metadata extracted from the HTML <title> tag
    - Other HTML <meta> tag data can be extracted
    - Parses and processes any links in the file
    - Links to other files in the collection are trapped and replaced by references to the document

Pasquale Savino - ISTI-CNR      Esempi - Il sistema Greenstone  
Dronetech Digitali      30

Pasquale Savino – ISTI-CNR      Esempi – Il sistema Greenstone  
Biblioteche Digitali      29

HTML Files

- ◆ **HTMLPlug Plug-In**
  - \*.htm
  - \*.html
  - \*.shtml
  - \*.shm
  - \*.asp
  - \*.php
  - \*.cgi

Pasquale Savino - ISTI-CNR      Esempi - Il sistema Greenstone  
Dronetech Digitali      30

Pasquale Savino – ISTI-CNR      Esempi – Il sistema Greenstone  
Bionotelle Digitali      29

Pasquale Savino – ISTI-CNR      Esempi – Il sistema Greenstone      Biblioteche Digitali      31

Pasquale Savino - ISTI-CNR  
Biblioteche Digitali  
Esempi - Il sistema Greenstone  
32

HTML Files

- ◆ ***file\_is\_url***
    - Optional switch within the HTML plug-in
    - Causes URL metadata to be inserted into each document, based on the file-name convention that is adopted by the mirroring package. The collection uses this metadata to allow readers to refer to the original source material rather than a local copy

Pasquale Savino – ISTI-CNR      Bioteche Digitali  
Esempi – Il sistema Greenstone      33

**Microsoft Word Files**

- ◆ WORDPlug Plug-In
    - \*.doc
  - ◆ Imports Microsoft Word documents
  - ◆ Greenstone uses independent programs to convert Word files to HTML
    - Many variants on the Word format
    - Older Word formats use a simple text string extraction

Pasquale Savino - ISTI-CNR      Esempi – Il sistema Greenstone      Bioteche Digitali      34

PDF Files

- ◆ PDFPlug Plug-in
    - \*.pdf
  - ◆ Imports PDF Files
  - ◆ Adobe's Portable Document Format
  - ◆ Greenstone uses independent programs to convert PDF files to HTML
  - ◆ PSPlug Plug-in
    - \*.ps
  - ◆ Imports PostScript Files
  - ◆ Works best when a standard conversion program is already installed on the computer
  - ◆ Uses simple text extraction algorithm if no conversion program is present

Pasquale Savino - ISTI-CNR      Esempi – Il sistema Greenstone      Bioteche Digitali      34

## PostScript Files

- ◆ PSPlug Plug-In
    - \*.ps
  - ◆ Imports PostScript Files
  - ◆ Works best when a standard already installed on the computer
  - ◆ Uses simple text extraction program is present

Pasquale Savino - ISTI-CNR      Esempi – Il sistema Greenstone      Bioteche Digitali      34

Email Files

- ◆ **EMAILPlug**
    - \*.email
  - ◆ **Imports files containing email**
    - Each source is checked for e-mail contents
  - ◆ **Extracts metadata:**
    - Subject
    - To
    - From
    - Date
  - ◆ **Deals with common formats**
    - Netscape, Eudora, Unix mail readers

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

Pasquale Savino - ISTI-CNR      Esempi - II sistema Greenstone      Biblioteche Digitali      38

**Compressed & Archived Files**

- ◆ ZIPPlug Plug-In
    - \*.zip
    - \*.tar
    - .gz
    - \*.z
    - \*.tgz
    - \*.bz
  - ◆ Relies on standard utility programs being present

- ◆ **Deals with common formats**
  - Netscape, Eudora, Unix mail readers

## | Classifiers

- ◆ Gestiscono strutture per il browsing della collezione
  - ◆ Vengono specificati nel Collection Configuration File
  - ◆ Per ogni classifier vi è una linea del tipo
    - `classify nome_classifier opzioni`
  - ◆ I programmati possano scrivere nuovi classifiers per creare nuove strutture di browsing

## Classifiers

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

Pasquale Savino - ISTI-CNR      Esempi - II sistema Greenstone      Biblioteche Digitali      38

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

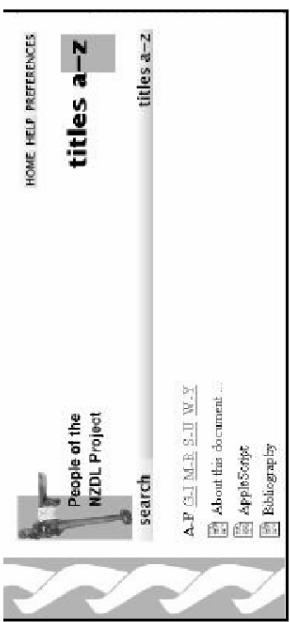
Pasquale Savino – ISTI-CNR  
Biblioteche Digitali  
Esempi – Il sistema Greenstone

Esempi di Classifier [1/4]

## Esempi di Classifier [2/4]

## ◆ AZList classifier

- Crea una lista ordinata alfabeticamente di elementi
  - Ad es. Classify AZList -metadata Title



Pasquale Savino - STI-CNR

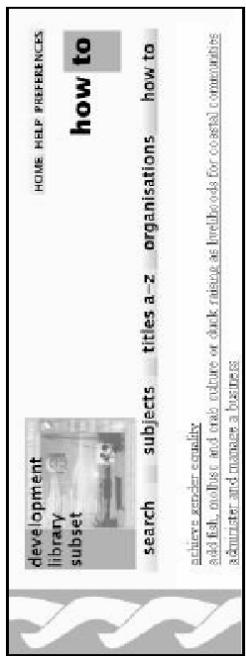
E-commerce - Il sistema Greenstone

41

42

## ◆ List classifier

- Crea una lista ordinata di elementi e li visualizza senza alcun ordine specifico
  - Ad es. `classify List -metadata Howto`

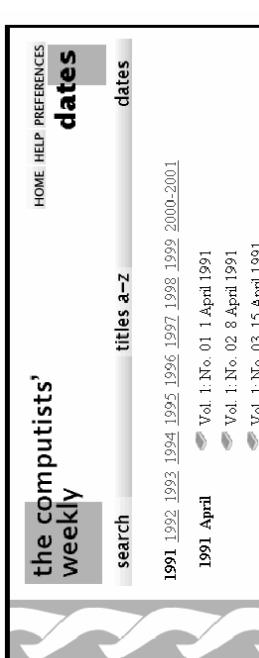


42

## Esempi di Classifier [3/4]

## ◆ DateList classifier

- Crea una lista ordinata di elementi data
  - Ad es. classify DateList -metadata date



Pasquale Savino - STI-CNR

E-commerce - Il sistema Greenstone

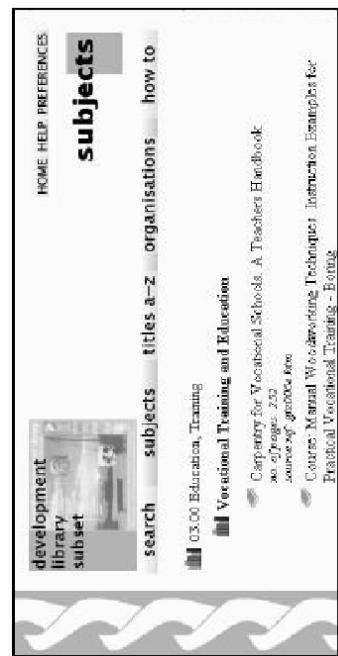
41

42

## Esempi di Classifier [4/4]

## ◆ Classifier gerarchici

- Creano classificazioni gerarchiche e sono utili per la classificazione di soggetti ed organizzazioni
  - Ad es. classify Hierarchy -hf file sub .txt - metadata Subject -sort Title



42

# I classifiers

- ◆ **Informazioni sui classifiers si possono avere digitando dalla linea comandi**
  - perl - S classinfo.pl nome\_classifier

Hierarchy	<i>hfile</i>	Hierarchical classification
	<i>metadata</i>	Classification file
	<i>sort</i>	Metadata element to test against <i>hfile</i> identifier
	<i>buttonname</i>	Metadata element used to sort documents within leaves (defaults to <i>Title</i> )
		Name of the button used to access this classifier (defaults to value of metadata argument)
<i>List</i>	<i>metadata</i>	Alphabetic list of documents
	<i>buttonname</i>	Include documents containing this metadata element
		Name of button used to access this classifier (defaults to value of metadata argument)
<i>SectionList</i>		List of sections in documents
<i>AZList</i>	<i>metadata</i>	List of documents split into alphabetical ranges
	<i>buttonname</i>	Include all documents containing this metadata element
		Name of button used to access this classifier (defaults to value of metadata argument)
		Like <i>AZList</i> but includes every section of the document
<i>AZSectionList</i>	<i>metadata</i>	Similar to <i>AZList</i> but sorted by date
	<i>buttonname</i>	Like <i>AZList</i> but sorted by date
		Sections
<i>DateList</i>		Esempi – Il sistema Greenstone
		Pasquale Savino – ISTI-CNR

45

46

# Searching Involves Indexes

- ◆ **Searching is provided by indexes built from different parts of the documents**
  - Entire documents
  - Paragraphs
  - Titles
  - Sections
  - Section headings
  - Figure captions

# Indexes

- ◆ **Indexes can be created automatically using**
  - Documents
  - Supporting files
- ◆ **Indexes can be rebuilt automatically**
  - New document in the same format becomes available
  - Process can awake, check for new material, and rebuild the indexes



## Introduzione

- ◆ Le pagine web visualizzate da Greenstone non sono persistenti ma vengono generate
- ◆ Le modalità di visualizzazione sono controllate dal comando “format” del Collection Configuration File
- ◆ Elementi della pagina controllabili
  - Item della pagina che presentano i documenti
  - Liste prodotte dai classifiers e risultati delle ricerche

## Come formattare l'output

Biblioteche Digitali  
Esempi – Il sistema Greenstone  
53

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

## Visualizzazione degli item nella pagina

*format DocumentImages true/false*  
*format DocumentHeadingFormatting*

*format DocumentContent true/false*

*format DocumentButtons string*

*format DocumentTextFormatting*

*format DocumentArrowsBottom true/false*

*format DocumentUseHTML true/false*

*format DocumentUseImage true/false*

<center><table width=37></tr><td>/Text</td></tr></table></center>

Display next/previous section arrows at bottom of document page (default true).

If true, each document is displayed inside a separate frame. The Preferences page will also change slightly, adding options applicable to a collection of HTML documents, including the ability to go directly to the original source document (anywhere on the Web) rather than to the Greenstone copy.

## Come formattare le liste

- ◆ Format lista - parte comandi
  - La prima parte (list) è obbligatoria ed identifica le liste alle quali applicare i comandi di formattazione
  - Search è la lista generata da una ricerca, mentre CL1, CL2, ... sono le liste generate dal primo, secondo, ... classificatore
  - La seconda parte (parte) è opzionale e specifica a quale parte della lista i comandi vanno applicati (HList, VList, DateList)
- ➔ Ad es. format CL4V1 list si applica a tutte le VList in CL4

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

56

## Come formattare le liste

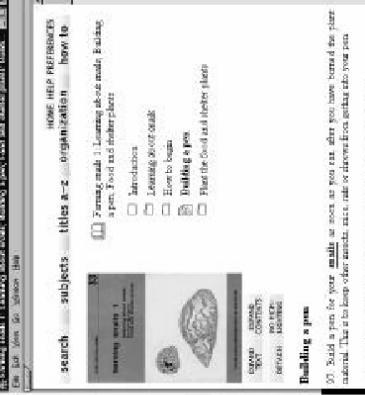
- ◆ Comandi è una stringa che specifica come formattare la lista

- ◆ Può contenere codice HTML, metadati ed i seguenti elementi

[Text]      The document's text  
[link]...    The HTML to link to the document itself  
[icon]     An appropriate icon (e.g. the little text icon in a Search Results string)  
[num]     The document number (useful for debugging).  
[metadata-name]    The value of this metadata element for the document, e.g. Title

Biblioteche Digitali  
Esempi – Il sistema Greenstone 57

Pasquale Savino – ISTI-CNR  
Biblioteche Digitali  
Esempi – Il sistema Greenstone 58



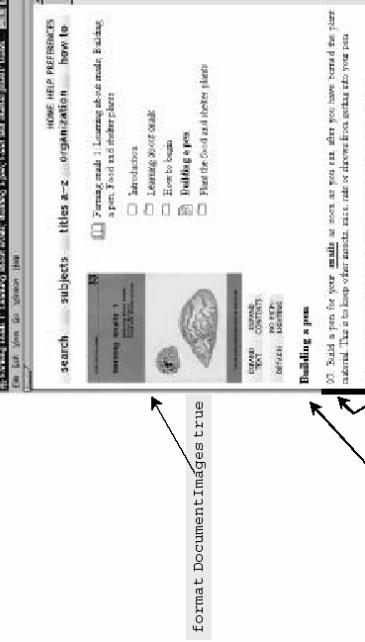
## Un esempio [3/4]

Howto classifier. È il quarto classifier (CL4)  
È un List classifier che genera una lista di titoli di documenti

```
1 classify Hierarchy -hfile sub.txt -metadata Subject -sort Title
2 classify AZList -metadata Title
3 classify Hierarchy -hfile org.txt -metadata Organisation -sort Title
4 classify List -metadata Howto
5 format SearchVList "<td align=top [<link>][icon]</td><td>{[If]
  {[parent (All : ) :Title], [parent (All : ) :Title] :}
  [<link>][Title]}</td></td>""
6 format CL4VList "<br>[Title]<br>[Howto]<br>""
7 format DocumentImages true
8 format DocumentText "<h3>[Title]</h3><br><br>[Text]""
9 format DocumentText
```

Comando di formattazione di CL4  
Gli elementi figlio degli elementi top-level sono visualizzati come una VList  
Ogni elemento si trova su una nuova linea e contiene il testo del campo Howto collegato al documento

Biblioteche Digitali  
Esempi – Il sistema Greenstone 59



## Un esempio [1/4]

- ◆ Esempio di classifiers e format commands della demo collection

```
1 classify Hierarchy -hfile sub.txt -metadata Subject -sort Title
2 classify AZList -metadata Title
3 classify Hierarchy -hfile org.txt -metadata Organisation -sort Title
4 classify List -metadata Howto
5 format SearchVList "<cd align=top [<link>][icon]</td><td>{[If]
  {[parent (All : ) :Title], [parent (All : ) :Title] :}
  [<link>][Title]}</td></td>""
6 format CL4VList "<br>[Title]<br>[Howto]<br>""
7 format DocumentImages true
8 format DocumentText "<h3>[Title]</h3><br><br>[Text]""
9 format DocumentText
```

Biblioteche Digitali  
Esempi – Il sistema Greenstone 60

# Un esempio [4/4]

The screenshot shows a search interface with a search bar containing 'pen'. The results list includes:

- Penazzaudi 1. Drawing a pen, Case illustrating Felt tip pen
- Penazzaudi 2. Drawing a pen, Case illustrating Felt tip pen
- Penazzaudi 3. Drawing a pen, Case illustrating Felt tip pen
- Penazzaudi 4. Drawing a pen, Case illustrating Felt tip pen
- Penazzaudi 5. Drawing a pen, Case illustrating Felt tip pen
- Penazzaudi 6. Drawing a pen, Case illustrating Felt tip pen
- Penazzaudi 7. Drawing a pen, Case illustrating Felt tip pen
- Penazzaudi 8. Drawing a pen, Case illustrating Felt tip pen
- Penazzaudi 9. Drawing a pen, Case illustrating Felt tip pen
- Penazzaudi 10. Drawing a pen, Case illustrating Felt tip pen

A callout box points to the first result with the text: "Lavorare su un matita o una matita".

Below the search bar, there is a link labeled "[link] [icon] [/link]".

On the left, a sidebar lists:

- 1 Classify Hie
- 2 Classify AZI [link] [Title] [/link]
- 3 Classify Hie
- 4 Classify List
- metadata Howto

On the right, a code editor shows the following code:

```
1 format CL4Vlist
2 format DocumentImages
3 format DocumentText
4 format SearchList
5 format SearchList
6 format SearchList
7 format CL4Vlist
8 format DocumentImages
9 format DocumentText
10 format DocumentText
```

The code includes annotations:

- An arrow points from the sidebar item "1 Classify Hie" to the line "format CL4Vlist".
- An arrow points from the sidebar item "2 Classify AZI [link] [Title] [/link]" to the line "format DocumentImages".
- An arrow points from the sidebar item "4 Classify List" to the line "format SearchList".
- An arrow points from the sidebar item "-metadata Howto" to the line "format DocumentText".