IMDI Browser

Version 3.1

IMDI Browser: Version 3.1
Version 3.1

Table of Contents

Introduction

The IMDI (ISLE Metadata Initiative) Browser was developed at the Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands. It is used to manage metadata and data files. As such, it complements the other IMDI tools, notably the IMDI Editor.

At the moment, the following IMDI tools exist (tools and manuals can be downloaded from http://www.lat-mpi.eu/tools/imdi):

- IMDI (Metadata) Editor: a tool for creating metadata descriptions, both for so-called sessions or resource bundles and for subcorpus nodes;
- IMDI Metadata Search Tool: a tool for searching through metadata information;
- IMDI CV-Editor: a tool for creating lists of controlled vocabulary items;

The IMDI Browser supports the following features:

- displaying and accessing data and metadata through a hierarchical tree structure and a world map;
- direct access to media, annotation and metadata files;
- searching of metadata files (by means of the IMDI Metadata Search Tool).

This manual explains and exemplifies the features of the IMDI Browser. It is organized around the following four chapters:

1. Basic information about the IMDI Browser (Chapter 1).
2. Displaying and accessing data (Chapter 2).
3. Searching data (IMDI Metadata Search Tool) (Chapter 3).
4. Changing files (Chapter 4).

Note

This manual contains a number of references to corpora housed at the Max Planck Institute for Psycholinguistics (MPI). It uses these corpora as a means to illustrate the functionality of the IMDI Browser. Please note that any remark (e.g., about the structure of the corpora, their availability, the contact person etc.) only pertain to these corpora. Corpora stored at other archives or institutes follow different rules.

1. Notation Conventions

The following notation conventions are used:

- Menu items, icons and screen displays are written in the font sans-serif.
- (Shortcut) keys are written in SMALL CAPS.

Note

Information on troubleshooting is printed like this

Chapter 1. Basic information

This section of the manual introduces you to the basic concepts and possibilities of the IMDI Browser. It introduces the concepts of corpora (Section 1.1) and of sessions or resource bundles[1] (Section 1.2). For illustration purposes, it makes use of the corpora stored at the Max Planck Institute for Psycholinguistics; for corpora stored at other institutes, different rules may apply.

1.1. Corpora

By default, the IMDI Browser gives access to all corpora that are stored at the Max Planck Institute for Psycholinguistics (MPI), Nijmegen, The Netherlands.

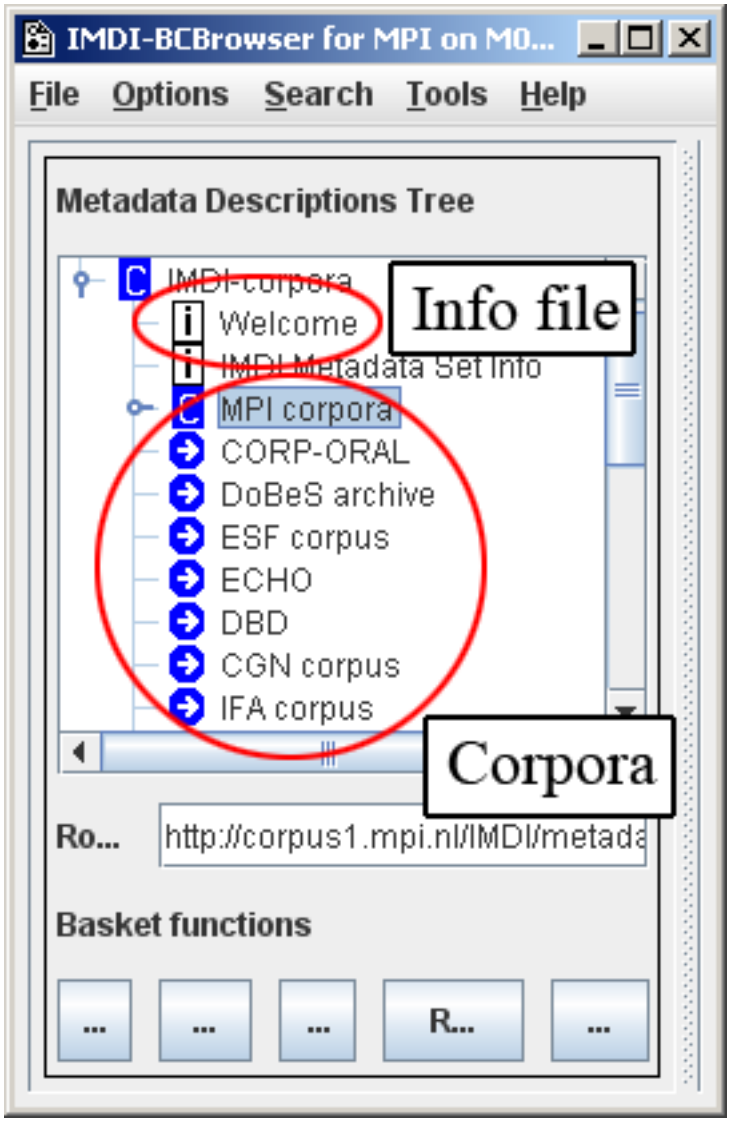


Figure 1.1. Corpora

In addition – or alternatively – the Browser can display corpora created by you or by other archiving initiatives (see Section 2.1.1).

Currently, some of the corpora housed at the MPI are:

[1] Please note: throughout this manual, the terms ‘session’ and ‘resource bundle’ are used interchangeably (see also Section 1.2).

- MPI corpora: Corpora collected by researchers affiliated with the Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands.
- DoBeS archive: Corpora collected by researchers affiliated with the DoBeS (Dokumentation Bedrohter Sprachen) project funded by the Volkswagen foundation.
- ESF corpus: European Science Foundation Second Language Acquisition Database.
- CGN: Corpus Gesproken Nederlands.[2]
- IFA corpus: Corpora collected by researchers affiliated with the Instituut voor Fonetische Wetenschappen (Institute of Phonetic Sciences), Amsterdam, The Netherlands.

Note

Depending on the version of the IMDI Browser, it may not always be possible to access all corpora.

Each corpus contains further subcorpora collected by individual researchers or project teams. The internal structure of each subcorpus varies according to the purposes and needs of the project. You can access information about their content and structure through the info files displayed in the IMDI Browser (see Section 2.3.4.3).

Each corpus contains metadata, media, annotation and info files (see Section 2.3.4). These files can be accessed through the following two mechanisms (but please note that data files may not always be accessible to the general public; see Section 1.2):

1. By navigating through the hierarchical tree structure (see Section 2.3.1).

   Each corpus is organized hierarchically in the form of tree structures (see Figure 1.2).

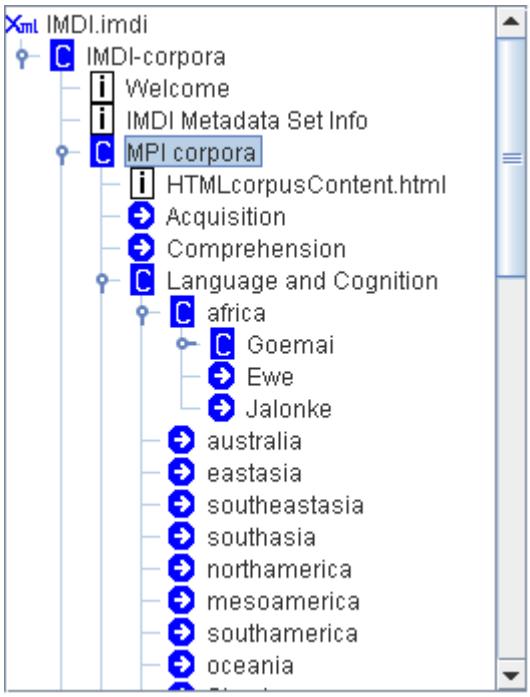


Figure 1.2. Corpus Tree Structure

   Tree structures consist of nodes that group files together on the basis of, e.g., the geographical region, the discourse genre, the sex or age of the speaker, the dialect of the speaker, the target/source language etc.

All nodes are displayed in the IMDI Browser. Double-click on a node to access the next level in the hierarchy.

At the lowest level of the hierarchy, the actual metadata, media, annotation and info files are displayed.

Note

The tree structure does *not* display the physical location of files. This means that the same file can easily be displayed in different parts of the corpus, e.g., the same file may be displayed under the nodes ‘folktale’, ‘male speaker’, and ‘age-group 20 to 30 years’.

2. By searching the metadata files (see Chapter 3).

   All resources or data files are accompanied by metadata files, i.e., files that give information about the data. These metadata files are searchable. For example, as illustrated below, you could search for all texts by female speakers above 60 years of age. The IMDI Browser displays the search results and allows you direct access to the corresponding data files.

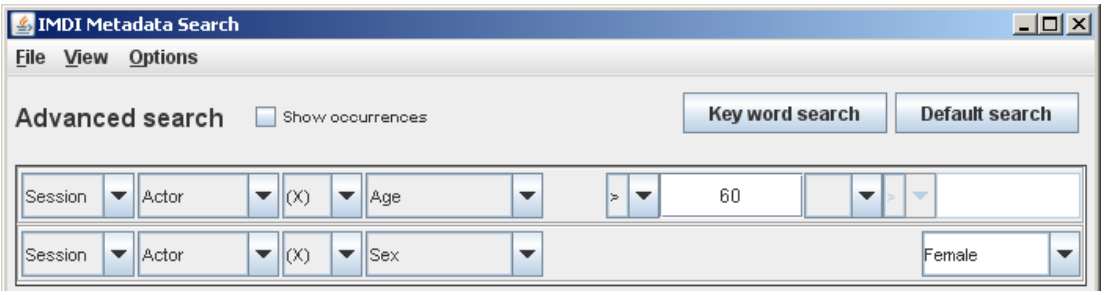


Figure 1.3. Metadata Search

1.2. Sessions or resource bundles

The lowest level of the corpus hierarchy contains the session or resource bundle. (These two terms are used interchangeably, ‘session’ being the preferred term for spoken corpora, and ‘resource bundle’ for written corpora.) A session or resource bundle corresponds to a meaningful unit of analysis, usually to a piece of data having the same overall content, the same set of participants, and the same location and time, e.g., one elicitation session on topic X, or one folktale, or one ‘matching game’, or one conversation between several speakers.

A session or resource bundle contains different files. The following four types of files exist:

1. Metadata files (see Section 2.3.4.1).

   These files contain information *about* the session, e.g., its date and location, its content and its participants. They are of the IMDI Editor format.

2. Media files (see Section 2.3.4.2).

   These files contain the audio or video recordings. They are usually digitized in one of the following formats: MPEG (`*.mpg`), Cinepak-Quicktime-Movies (`*.mov`), WAVE (`*.wav`).

3. Annotation files (see Section 2.3.4.2).

   These files contain the transcripts, codings and annotations. Their format varies (e.g., ELAN, Media Tagger, Shoebox, CHAT, etc.)

4. Info files (see Section 2.3.4.3).

These files contain further background information on specific topics. They are in PDF or HTML formats.

With a few exceptions, metadata and info files are publicly accessible. This policy enables researchers to search metadata and to thereby gain an overview of the available material.

Annotation and media files are handled differently in the different corpora: some corpora allow for general access (e.g., the CGN corpus), while others do not (e.g., the MPI corpora). In the latter case, if you have not been explicitly granted access by the responsible researcher(s), you will not be able to access the files. To gain access, ask the responsible researcher(s) for permission. Contact details are displayed in each corpus, e.g.:

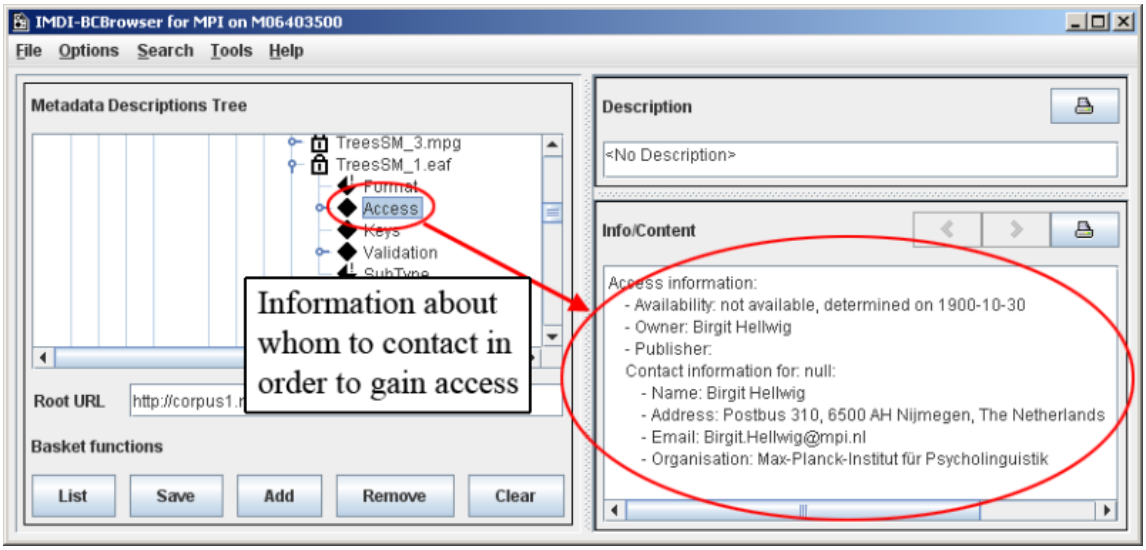


Figure 1.4. Contact Details

Responsible researcher(s), by contrast, have unrestricted access to all files. This includes the following possibilities:

- Access to metadata and info files, including (a) all suppressed information and (b) the possibility to update the information. See Section 2.3.4.1 and Section 2.3.4.3.
- Access to media files. See Section 2.3.4.2.
- Access to annotation files, including the possibility to continuously update them. See Section 2.3.4.2.

Chapter 2. Displaying and accessing data

Starting the IMDI Browser opens up the window IMDI-BCBrowser for MPI (referred to in this manual as the IMDI Browser window). In the IMDI Browser window, you can view and access the corpora, i.e., you can read information about the available data, access the files, and initiate searches. The IMDI Browser window contains the following information:

1. Menu items (see Section 2.1)
2. Bookmarks panel (see Section 2.2)
3. Metadata Descriptions Tree panel (see Section 2.3)
4. Info/Content panel (see Section 2.4)
5. Description panel (see Section 2.5)

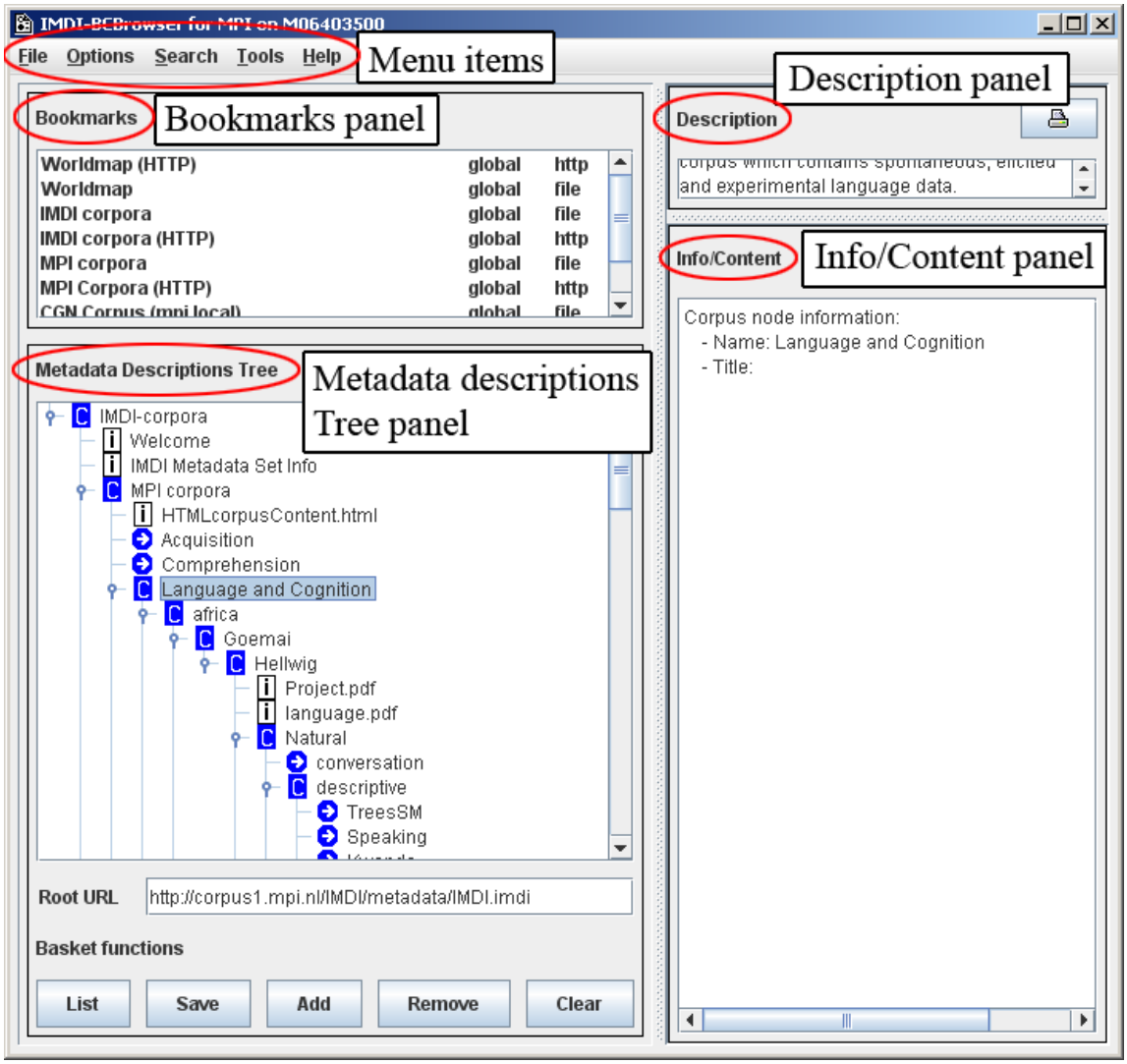


Figure 2.1. IMDI-BCBrowser for MPI

2.1. Menu items

The following five menu items are available: File (see Section 2.1.1), Options (see Section 2.1.2), Search (see Section 2.1.3), Tools, Help (see Section 2.1.5).

To access any menu item, do one of the following:

1. Click on that item.

2. Or use the shortcut key ALT plus the underlined letter(s) (e.g., to access the item File, press the key ALT+F).

2.1.1. File menu

The File menu supports the following two options:

Figure 2.2. File Menu

- Open File

  Click on Open File to open a *corpus* file, i.e., an `*.imdi` corpus file that has been created with the IMDI Editor and that contains information about the corpus.

  The Open dialog box appears, allowing you to browse to the folder that contains the `*.imdi` corpus file. In the dialog box, click on Open, and the IMDI Browser window will display the selected corpus node. E.g.:

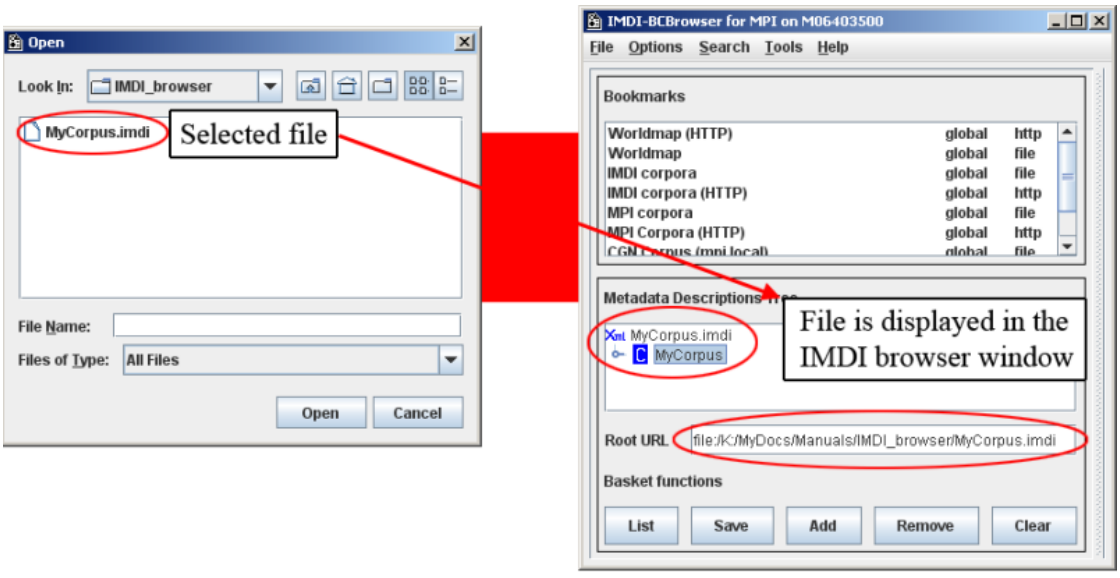


Figure 2.3. Open File

- Exit

Click on Exit to exit the IMDI Browser window.

Note that the Exit item turns into a Close item when you open a second IMDI Browser window, either through the option Clone Node (see Section 2.3.1) or through accessing your search results (see Section 3.4). If you click on Close, only the current IMDI Browser window will be closed.

2.1.2. Options menu

The Options menu supports the following six options:

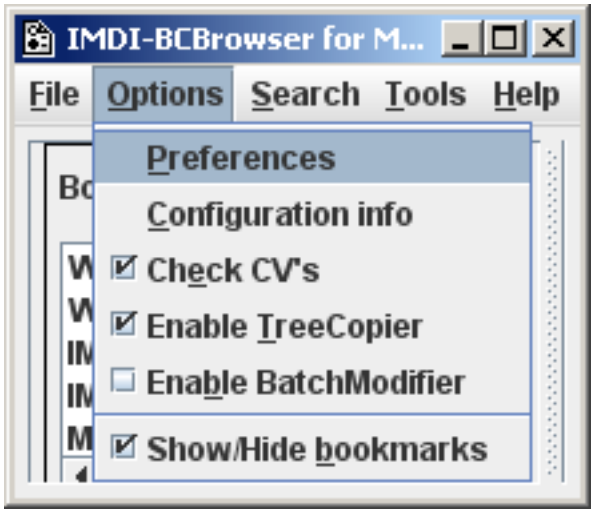


Figure 2.4. Options Menu

- Preferences

  Click on Preferences to select the update policy, fonts network setting and IMDI DBcreation settings. The Preferences dialog window appears, offering you the following two options:

  – Update Policy

    Click on Update Policy to determine the policy for updating controlled vocabularies stored in the cache. The following policies are available:

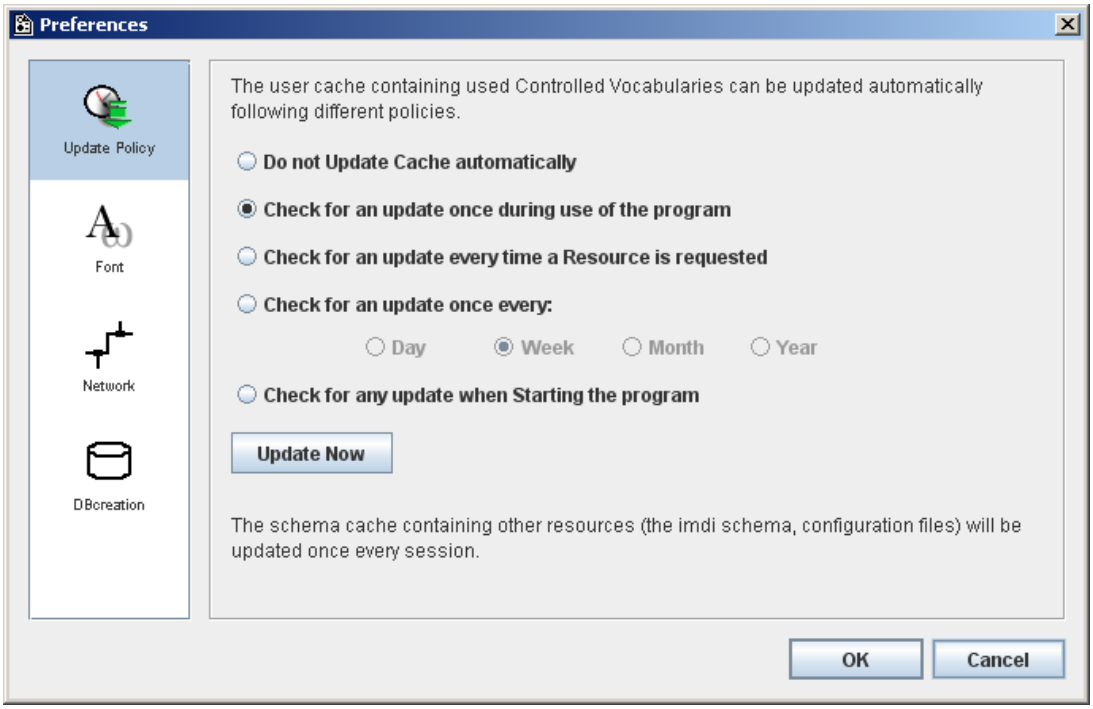


Figure 2.5. Preferences

Note

The IMDI tools make use of so-called “controlled vocabularies” (i.e., lists of values for a specific key) to ensure consistency across users. These vocabularies are stored on the Web, and a copy is stored in the cache of your desktop computer (in the folder “user-directory \ IMDI-TOOLS \ CVCACHE”). The Web version is changed sometimes, e.g., new values are added. However, the IMDI tools will continue to use the locally-stored version until the cache is updated again. By default, it is updated whenever you run the IMDI Browser. However, you can change this policy, and determine if and when it should be updated.

– Font

This option allows you to display information from metadata files that contains special character sets. To select a set and a font, do the following:

1. Under Unicode Block (in the left panel), select a unicode block by clicking on it. The available fonts for that block are displayed under Font (in the right panel).

2. Under Font, select a font by clicking on it.

3. Click on OK to implement the new font in the Browser.

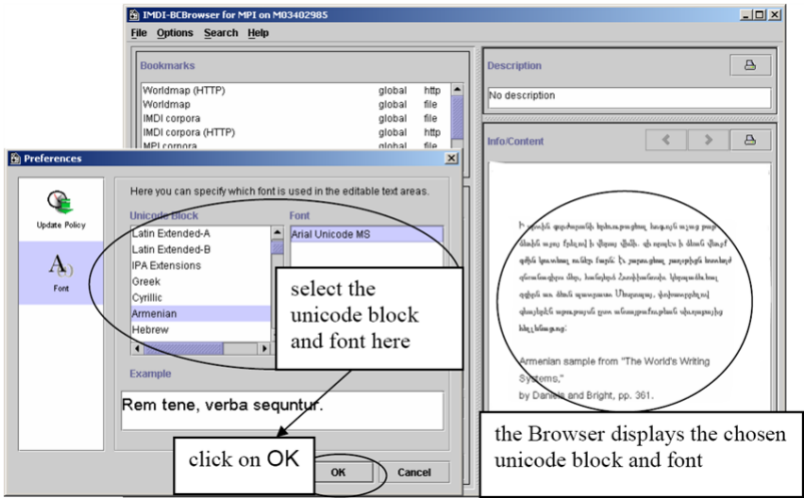


Figure . Font

– Network

If you want to access a corpus via the HTTP protocol, you might have to use a so-called proxy server. To do so, select the Network icon in the preferences dialog, check the use HTTP checkbox and provide the proxy sever's name and port number.

– DB Creation

In order to host a public corpus, it is necessary to setup a public database as well, so that users can perform search operations without having to create the search index files themselves. To alter the settings for such a server, have a look at the *DB Creation* icon in the Preferences dialog. Here you can edit the folllowing options:

—Enable DB Creation remote files: If this option is enabled, the `*.data` files (see Section 3.1.2) are stored inside the corpus file structure, instead of in the user's personal directory.

—Validate XML: Always check whether the files to be added to the search database are well-formed XML files. Takes some extra time.

—Include Descriptions: If true the information in the "free text" description elements is included in the DB. Default is false.

—SeparateKeyValueTables: Set this to false if the DB files concern a single corpus that systematically uses the same set of user defined keys. Default is true.

—Local Corpus Root: The part of the URL path seen by the IMDI brower that all IMDI files of the corpora have in common. This information is used to make the path names in the DB files relative.

—HTTP Corpus Root: The URL path seen by a www-browser that all IMDI files of the corpus have in common.

—UNIX (NFS) Local Corpus Root: The URL path seen by a (UNIX) NFS client that all IMDI files of the corpus have in common.

—Windows (SMB) Local Corpus Root: The URL path seen by a (Win) SMB client that all IMDI files of the corpus have in common.

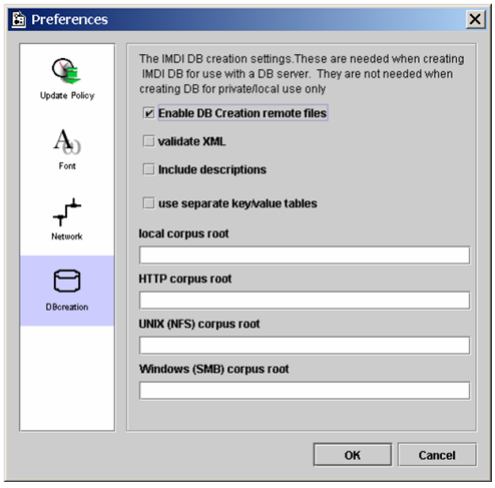
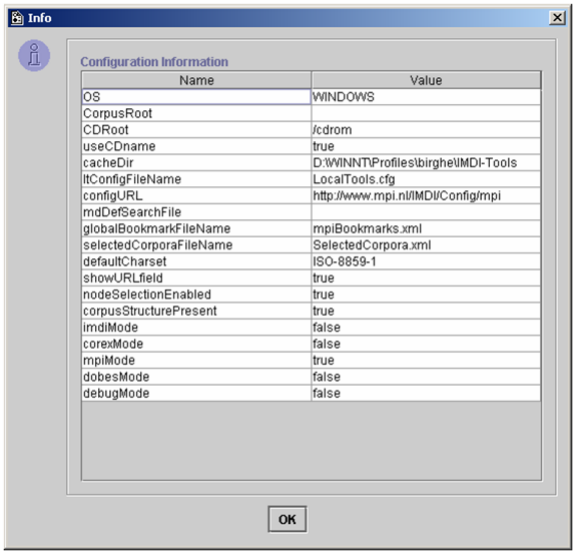


Figure 2.7. DB Creation

- Configuration Info

  Click on Configuration Info to view information on the configuration of the tool (you might need this information when reporting a bug to the developers), e.g.:



Name	Value
OS	WINDOWS
CorpusRoot	
CDRoot	/cdrom
useCDname	true
cacheDir	D:\WINNT\Profiles\birghe\IMDI-Tools
ltConfigFileName	LocalTools.cfg
configURL	http://www.mpi.nl/IMDI/Config/mpi
mdDefSearchFile	
globalBookmarkFileName	mpiBookmarks.xml
selectedCorporaFileName	SelectedCorpora.xml
defaultCharset	ISO-8859-1
showURLfield	true
nodeSelectionEnabled	true
corpusStructurePresent	true
imdiMode	false
corexMode	false
mpiMode	true
dobesMode	false
debugMode	false

Figure 2.8. Configuration Info

- Check CV's

  Click in the box to the left of Check CV's to enable the IMDI Browser to check if the values used in a metadata file are correct, i.e., are specified in the controlled vocabulary.

- Enable TreeCopier

  Check this option if you want to be able to download a corpus to your local harddrive.

Note

This option does not work if the server runs Apache and a Shibboleth security engine.

- Enable BatchModifier

  Check this option if you want to change files (see Chapter 4).

- Show/Hide bookmarks

  Click Show/Hide bookmarks to show or hide the Bookmarks panel in the IMDI Browser window.

2.1.3. Search menu

The Search menu supports the following option:

Figure 2.9. Search Menu

- Metadata Search

  Click on Metadata Search to access the IMDI Metadata Search tool (see Chapter 3).

2.1.4. Tools menu

The Tools menu supports the following option:

Figure 2.10. Options Menu

- Selected Sessions

  If there are sessions in the basket (see Section 2.3.3), clicking this option will display these sessions in the Info/Content Panel.

2.1.5. Help menu

The Help menu supports the following option:

Figure 2.11. Help Menu

- About

  Click on About to view the copyright and version information.

2.2. Bookmarks panel

In the Bookmarks panel, you can save shortcuts, i.e., “bookmarks”, to parts of the corpus. Such bookmarks have the advantage that you do not need to navigate through the entire corpus hierarchy in order to access that part of the corpus.

Depending on the version and the configuration of the IMDI Browser, a number of bookmarks are displayed by default. For example, the following default bookmarks give access to the corpora housed at the Max Planck Institute for Psycholinguistics:

- World Map (HTTP)

  Access (via the web server) to a world map that displays the locations of documented languages.

- World Map

  Access (via the local MPI network) to a world map that displays the locations of documented languages.

- IMDI Corpora

  Access (via the local MPI network) to the MPI, DoBeS, ESF, CGN and IFA corpora.

- IMDI Corpora (HTTP)

  Access (via the web server) to the MPI, DoBeS, ESF, CGN and IFA corpora.

- MPI Corpora

  Access (via the local MPI network) to the MPI corpora.

- MPI Corpora (HTTP)

  Access (via the web server) to the MPI corpora.

- CGN Corpus (mpi local)

  Access (via the local MPI network) to the CGN corpus.

- CGN Corpus (HTTP)

  Access (via the web server) to the CGN corpus.

- Search Results

  Access to your search results.

Whenever you double-click on a bookmark, the corresponding node is displayed in the Metadata Descriptions Tree panel (see Section 2.3).

In addition to the predefined bookmarks, you can create your own bookmarks. Do the following:

1. In the Metadata Descriptions Tree panel (see Section 2.3), navigate to the corpus or session node for which you want to create a bookmark.

2. Click on the node to select it. It will be highlighted in blue.

3. Click with the right mouse button on the node to open a pull-down menu.

4. Select Add to Bookmarks from the pull-down menu. The Input dialog box appears.

5. Specify a name for the bookmark, and click on OK. The new bookmark is added to the Bookmarks panel, and will be marked as personal.

E.g.:

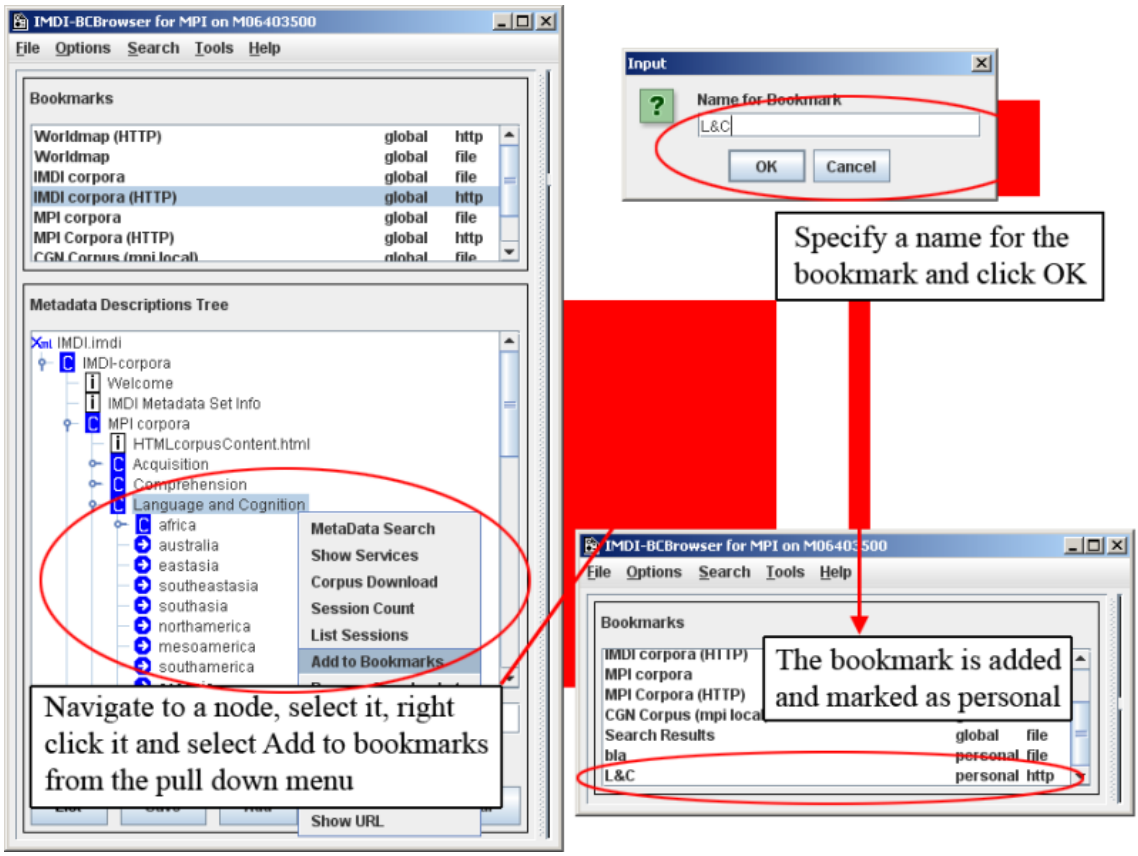


Figure 2.12. Adding a bookmark

The bookmarks are saved permanently by the IMDI Browser, i.e., they remain available every time you restart the Browser.

To remove a bookmark, do the following:

1. In the Bookmarks panel, click on the bookmark that you want to remove. It will be highlighted in blue.

2. Click with the right mouse button on the bookmark to open a pull-down menu.

3. Select Delete Bookmark from the pull-down menu. The bookmark is deleted without further warning

Note

You can only delete bookmarks created by yourself (i.e., those marked personal in the Bookmarks panel), but not the predefined bookmarks (i.e., those marked global).

2.3. The Metadata Descriptions Tree panel

The Metadata Descriptions Tree panel allows you to navigate through the corpus hierarchy. It serves the following three purposes:

1. The hierarchy of nodes allows you to easily navigate through the corpus to the session data (see Section 2.3.1 and Section 2.3.2).

2. Pull-down menus and the five buttons at the bottom of the panel allow you to select parts of the corpus for purposes of conducting searches (see Section 2.3.3).

3. The linking of metadata, media and annotation files at the session level allows you to immediately access all relevant data (see Section 2.3.4).

The following screenshot illustrates the Metadata Descriptions Tree panel.

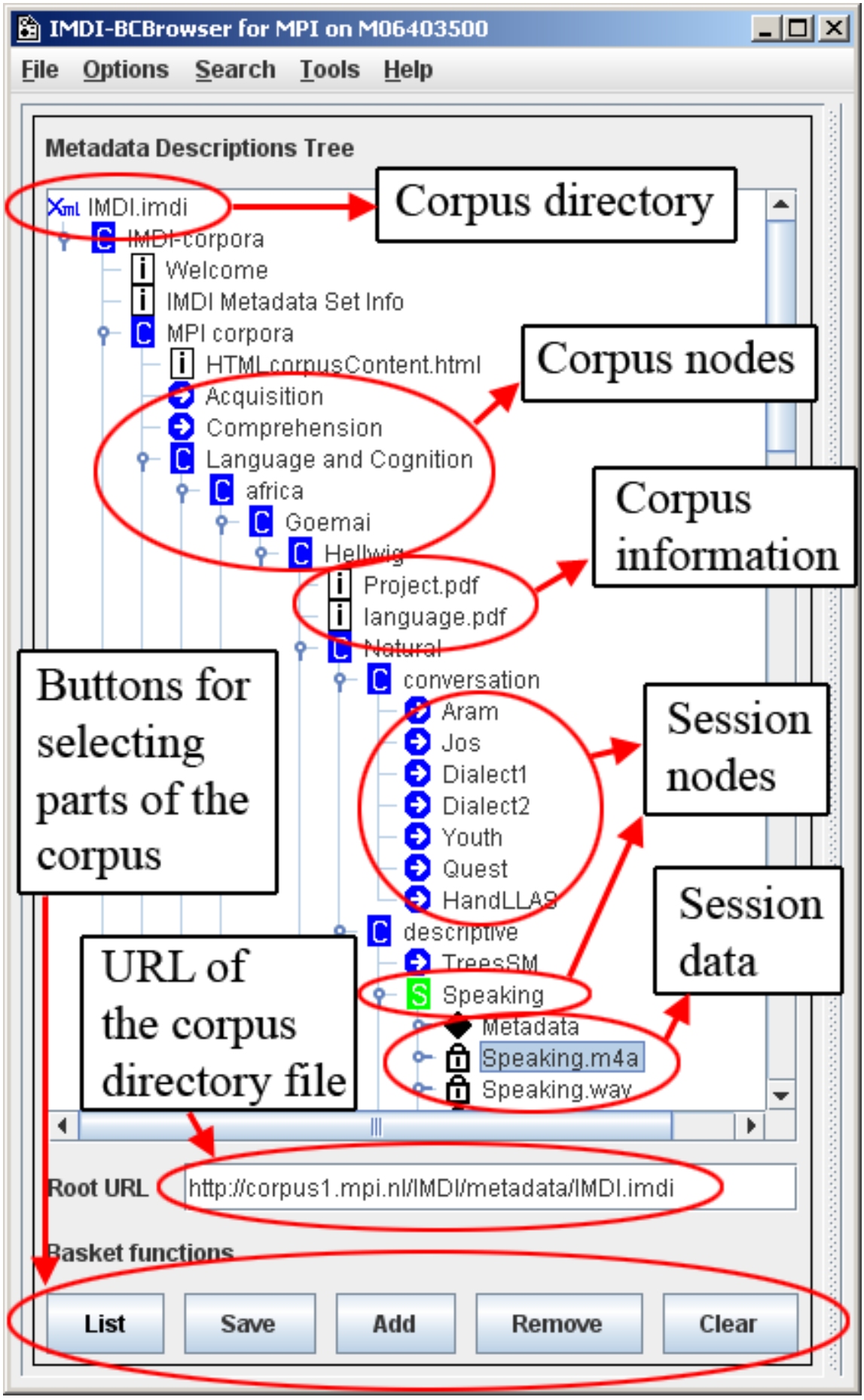


Figure 2.13. Metadata description tree panel

2.3.1. Navigating through the corpus (tree structure)

Whenever the IMDI Browser window opens up, the node IMDI Corpora is displayed in the Metadata Descriptions Tree panel.

If you want to access any other predefined part of the corpus, do the following:

1. Go to the Bookmarks panel (see Section 2.2).
2. Double-click on one of the bookmarks. The corresponding node is displayed in the Metadata Descriptions Tree panel.

In the Metadata Descriptions Tree panel, make use of the following options to navigate through the corpus:

- Double-click on any node to open it and display the next level in the hierarchy.

Note

The Metadata Descriptions Tree panel distinguishes between open and closed corpus and session nodes. These are represented through the following icons:

–  icon of an open corpus node

–  icon of an open session node

–  icon of a closed corpus/session

Some program commands do not work when a node is closed. If any of the commands do not seem to work, make sure that the node is open.

Note

Because of the large amount of data that is loaded, it may take some time until the IMDI Browser responds to your command and opens a node.

- Click on any closed node, and then point with the mouse to it to activate the tool-tip that displays brief information about the node.
- Click on any open node to select it. It will be highlighted in blue. The information relevant to the selected node is displayed in the Info/Content (see Section 2.4) and Description panels (see Section 2.5).
- Right-click on any selected (i.e., highlighted) node to open a pull-down menu that displays the available options. Click on any item in the pull-down menu to select the corresponding option.

  Depending on the kind of node, the following options are available:

  – Add to basket

    Adds the node to the list of nodes to be searched (see Section 2.3.3).

  – Add to Bookmarks

    Adds the node to the Bookmarks panel (see Section 2.2).

  – Bookmark Info

    Displays information about the bookmark in the Info/Content panel (see Section 2.4).

- Clone Node

  Opens a second IMDI Browser window that displays only the corresponding node.

- Create Service Files

  Creates a file that is needed for searching through the metadata (see Section 3.1.2).

- Delete Bookmark

  Deletes the node from the Bookmarks panel (see Section 2.2).

- Delete Local Service Files

  Deletes the locally-stored file that is needed for searching through the metadata (see Section 3.1.2).

- ELAN

  Opens the media file or the session annotation in ELAN (see Section 2.3.4.2).

- IMDI-BCEditor

  Opens the metadata file in the IMDI Editor (see Section 2.3.4.1).

- List Sessions

  Lists all sessions contained under this node in the Info/Content panel (see Section 2.4).

- MetaData Search

  Opens the IMDI Metadata Search window (see Chapter 3).

- Open with Acrobat Reader 5.0

  Opens the file with Acrobat Reader 5.0 (see Section 2.3.4.3).

- Open with Winamp

  Opens the file with Winamp (see Section 2.3.4.2).

- Remove

  Removes the node from the Metadata Descriptions Tree panel.

- Remove from basket

  Removes the node from the list of nodes to be searched (see Section 2.3.3).

- Save File Content

  Downloads the file in a non-compressed format.

- Session Count

  Displays the number of sessions contained under this node in the Info/Content panel (see Section 2.4).

- Show

  Displays the file content in the Info/Content panel (see Section 2.3.4.2 and Section 2.4).

- Show As HTML

Displays the file in an HTML format in the Info/Content panel (see Section 2.3.4.1 and Section 2.4).

– Show Content-Type

Displays the file format in the Description panel (see Section 2.5).

– Show Description

Displays the file description in the Description panel (see Section 2.5).

– Show File Content

Displays the file in an XML format in the Info/Content panel (see Section 2.4).

– Show format

Displays the file format in the Info/Content panel (see Section 2.4).

– Show Info

Displays the file content in the Info/Content panel (see Section 2.4).

– Show LR's

Displays the directory information for all files contained under this node in the Info/Content panel (see Section 2.4).

– Show Services

Displays the name and directory information of the file needed for searching through the metadata in the Info/Content panel (see Section 3.1).

– Show URL

Displays the directory information for the metadata file in the Description panel (see Section 2.5).

– Windows Media Player

Opens the media file in Windows Media Player (see Section 2.3.4.2).

Note

Depending on the configuration of your IMDI Browser, other tools may be available. Furthermore, under Windows, the default option for starting media files of specific types is also supported (e.g., if Windows is set to open `*.wav` files in Windows Media Player, the option Windows Media Player will be available).

Note

For some nodes, one of the options is marked with the symbol #. In this case, the corresponding option will start automatically whenever you double-click on the open node.

Note

Not all options are available for all nodes.

2.3.2. Navigating through the corpus (world map)

The IMDI Browser displays a world map with the locations of all documented languages. To access the world map, do the following:

1. In the Bookmarks panel double-click on the bookmark Worldmap.

   The info file `world.html` is displayed in the Metadata Descriptions Tree panel.

2. In the Metadata Descriptions Tree panel click on the info file `world.html`.

   The world map is displayed Info/Content panel.

3. To navigate, you have the following options:

   - Click on any part of the world map to enlarge the corresponding region.
   - Make use of the Backward and Forward buttons to move to previous and subsequent views.
   - Click on any language to open the corresponding node in the Metadata Descriptions Tree panel.

For example:

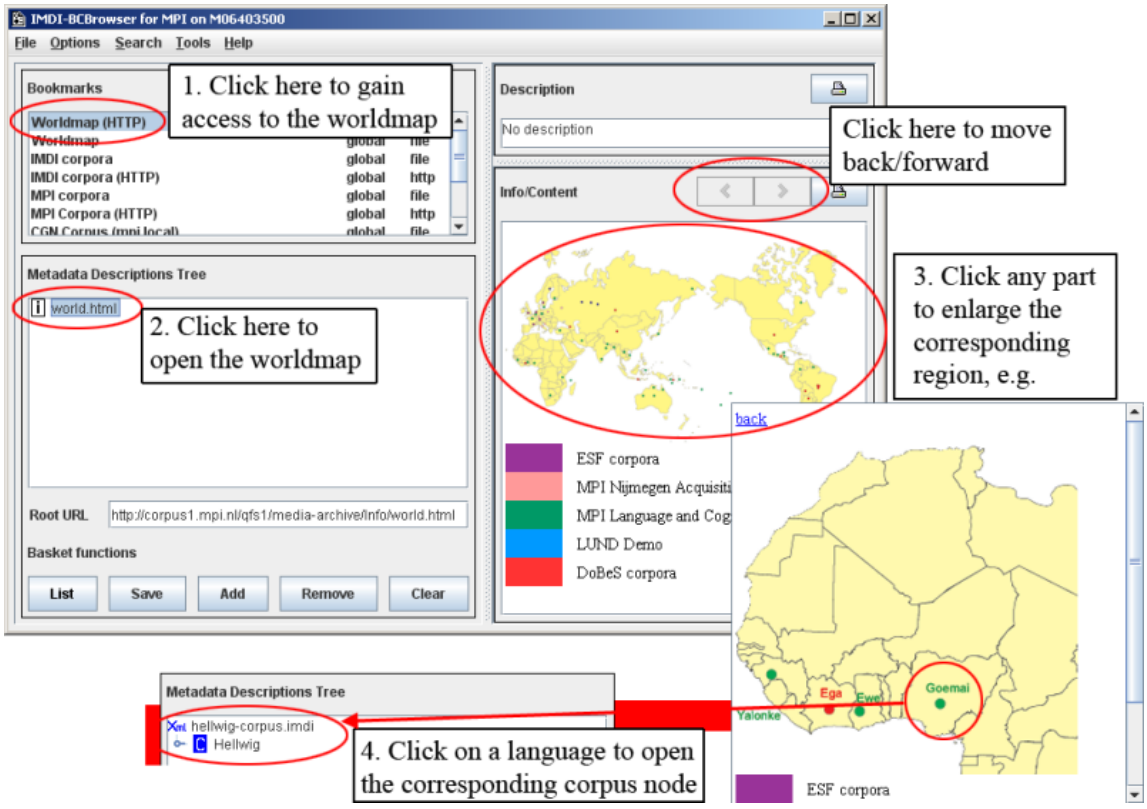


Figure 2.14. World map corpus navigation

2.3.3. Selecting parts of the corpus for purposes of searching (basket function)

The IMDI Browser allows you to search through the metadata files. By default, this search is done throughout an entire corpus (see Section 3.1), but it is possible to limit it to one (or several) selected corpus and/or session nodes, i.e., to limit it to nodes that have been put in the “basket”.

To put a corpus or session node into the basket, do the following:

1. In the Metadata Descriptions Tree panel, double-click on the node to open it.

2. Click on the open node to select it. It will be highlighted in blue.

3. Do one of the following:

   a. Click the Add button at the bottom of the Metadata Descriptions Tree panel.

   b. Or click with the right mouse button on the highlighted node; then select Add to basket from the pull-down menu.

   The icon of any selected node will change its color to gray, e.g.:

   -  non-selected node

   -  selected node

   Once an item is selected, the List button at the bottom of the panel will be highlighted in red.

Note

The red color was chosen to alert you to the fact that a list has been compiled. The existence of such a list has consequences for the search process: metadata search will only search the listed nodes, i.e., the nodes that were put into the basket (see Section 3.1). If the search process does not yield the expected results, please make sure that the list contains all relevant nodes. And if you want to search through an entire corpus (not just through the listed nodes), you have to delete the list again (see below).

4. Repeat this process to add other nodes to your selection.

Note

If you wish to search a locally stored corpus, please select only nodes that are listed below a corpus node for which there is a search file available. Such corpus nodes are visualized in the IMDI Browser by means of a white “C” on a black background (see Section 3.1). As indicated in Figure 2.15, you can select and search nodes contained under MPI corpora *or* under DoBeS archive. But you cannot select and search one node from MPI corpora *and* one from DoBeS archive – since they do not have a common search file.

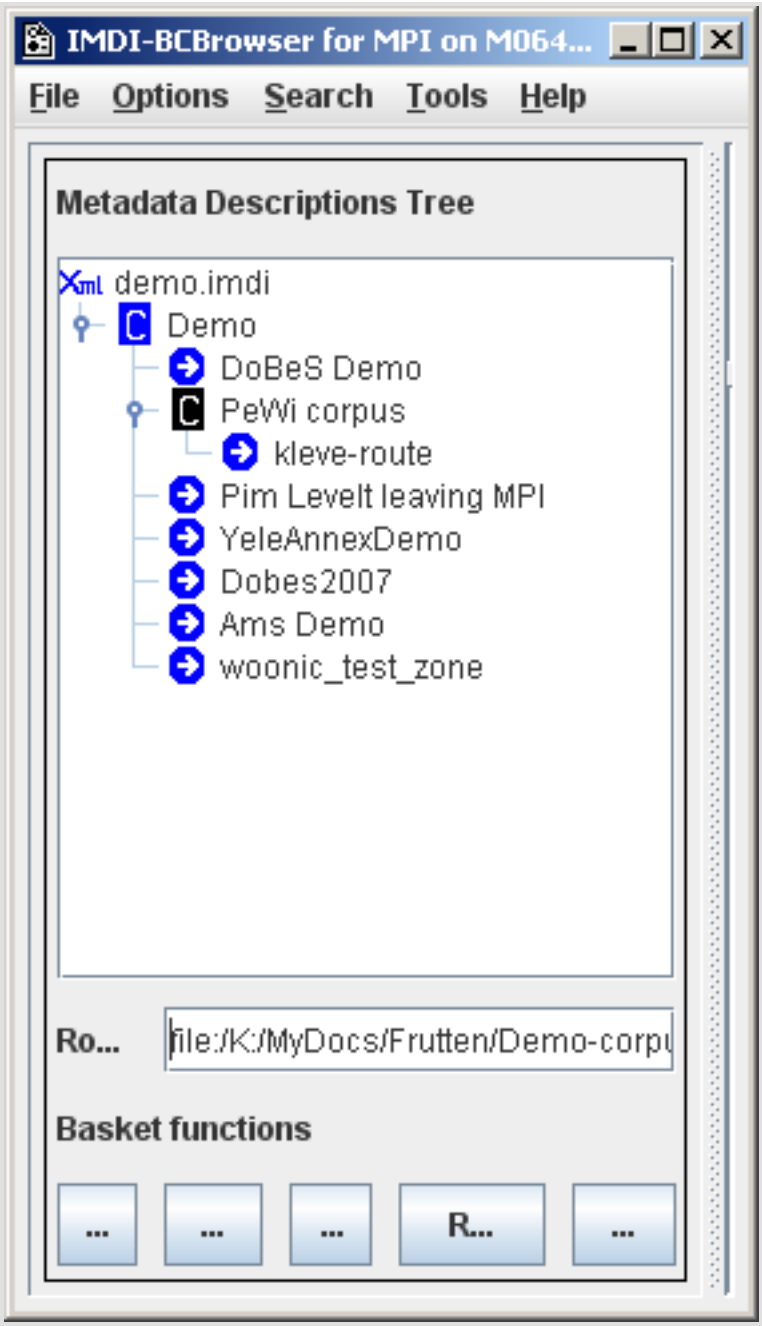


Figure 2.15. Corpora with a search file

To view a list of the selected nodes, click the List button at the bottom of the Metadata Descriptions Tree panel. It will display all selected nodes, e.g.:

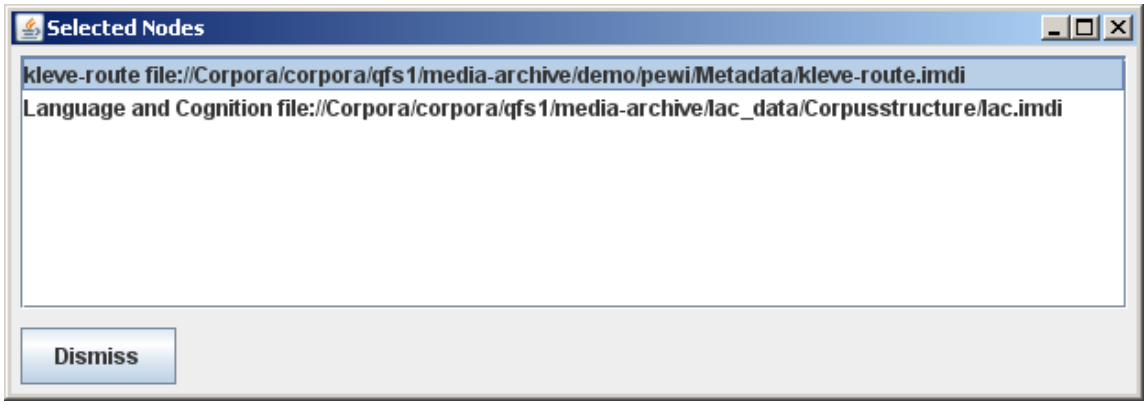


Figure 2.16. List of selected nodes

To remove a node from the list, do the following:

1. In the Metadata Descriptions Tree panel, double-click on the node to open it.
2. Click on the open node to select it. It will be highlighted in blue.
3. Do one of the following:
   a. Click the Remove button at the bottom of the Metadata Descriptions Tree panel.
   b. Or click with the right mouse button on the highlighted node; then select Remove from basket from the pull-down menu.
4. Repeat this process to remove other nodes from your selection.

You can remove all selected nodes from this list by clicking the Clear button at the bottom of the Metadata Descriptions Tree panel.

You can save the selected list for future uses. Click the Save button at the bottom of the Metadata Descriptions Tree panel. The following message informs you that your list has been saved:

Figure 2.17. Saving a list

Note

Once you have saved a selected list, you can only remove it by first clicking the Clear button (to remove all selected nodes) and then the Save button (to save the empty list).

When you are satisfied with your selection, you can initiate the search (see Section 3.2.3.4).

If you only want to perform a search in one subcorpus, you can also right click on this node and select MetaData Search from the context menu instead of adding it to the search basket:

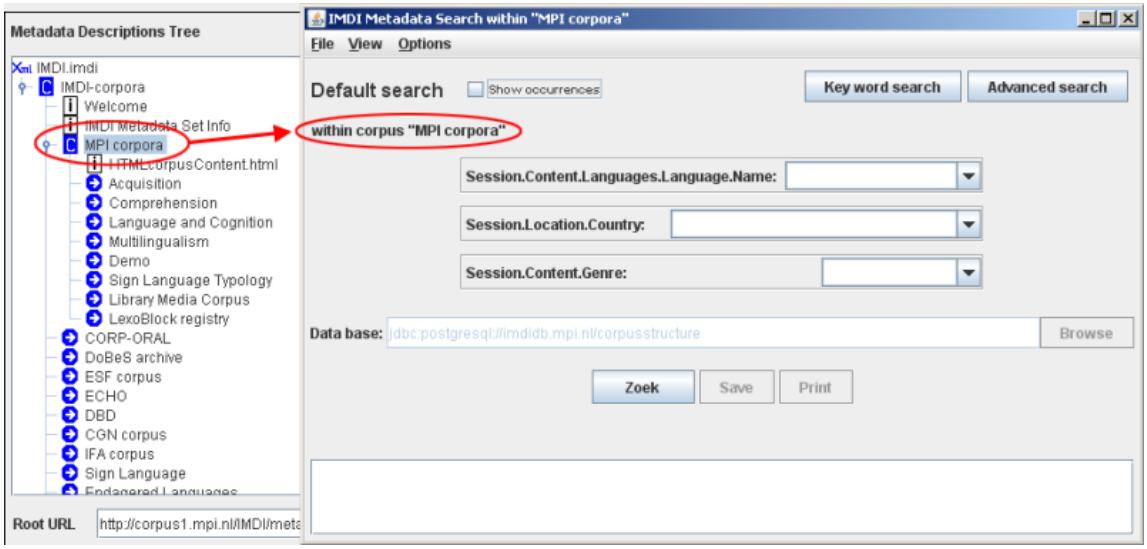


Figure 2.18. Metadata search in one corpus

2.3.4. Accessing data

The IMDI Browser contains links to files or to information from files. These links are symbolized through the following icons:

- ◆ link to information from metadata file(s) (i.e., information from IMDI files).
- m link to digitized media file(s), containing audio/video data.
- T link to transcription and annotation file(s).
- i link to info file(s) providing general background information.
- ⛔ link not available (either because the file does not yet exist, or because access is denied).

Click on any of these icons, and the IMDI Browser will display the relevant information.

2.3.4.1. Metadata information

Metadata information is displayed under each session node. It allows you to (a) read about the circumstances under which the session data was collected and (b) search for relevant data (see Chapter 3).

It contains the following kind of information:

- Information about the date and location.
- Information about the project within which the data was collected.
- Information about the person who collected the data.
- Information about the content.

- Information about the participant(s).
- Information about the source (i.e., the audio/video tape), the digitized media file(s) and the annotation file(s).
- Cross-references to other relevant sessions and publications.

To view the metadata information of a session, click on the corresponding metadata icon. The information is displayed in the Info/Content panel (see Section 2.4) or the Description panel (see Section 2.5).

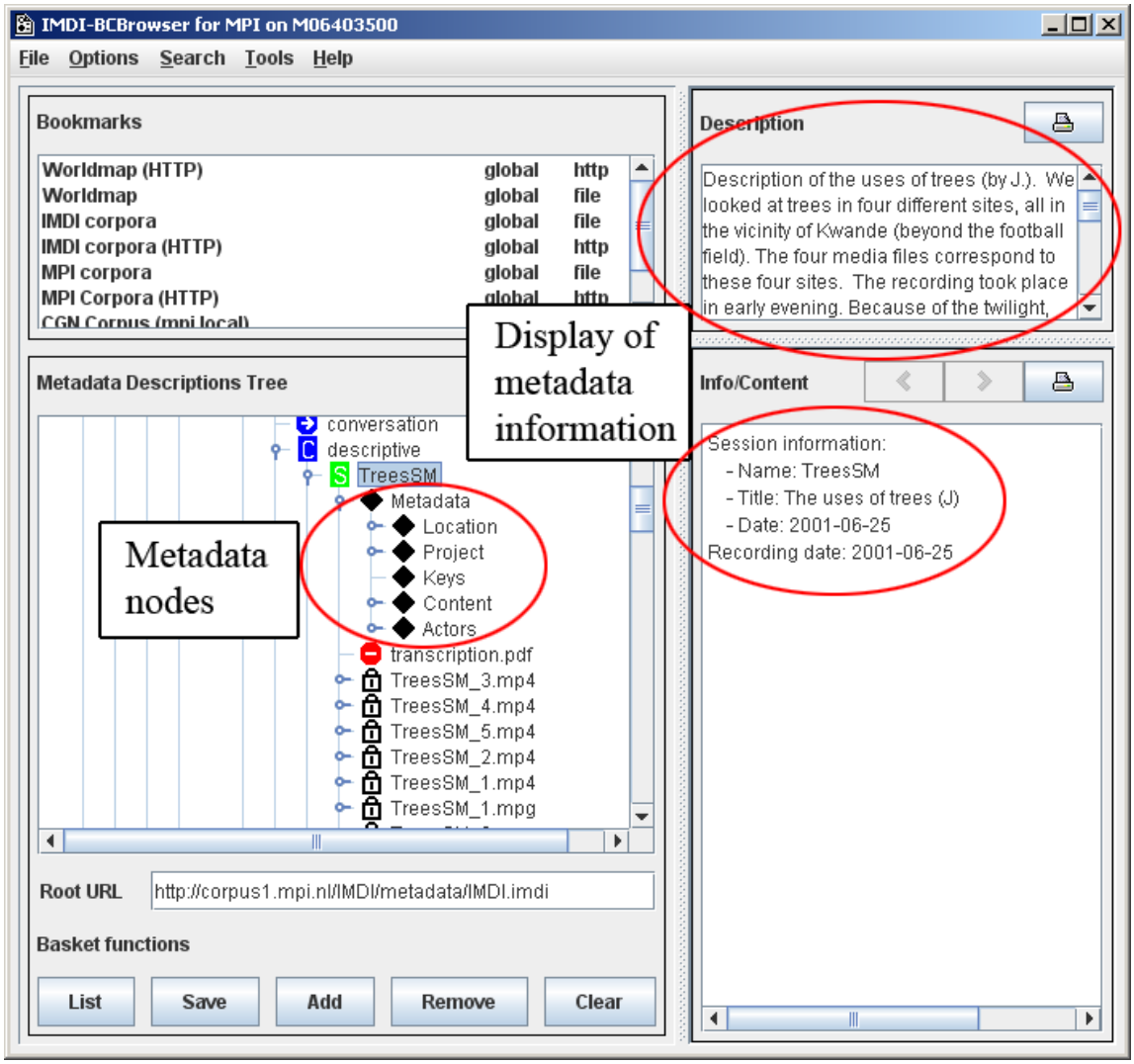


Figure 2.19. Metadata

Furthermore, you can access all metadata information through the session node. Double-click on a session node to open it, then click on it to select it, and right-click to open a pull-down menu. The pull-down menu offers you the following options:

- Choose Show As HTML from the menu. The metadata information is displayed in an HTML format in the Info/Content panel, e.g.:

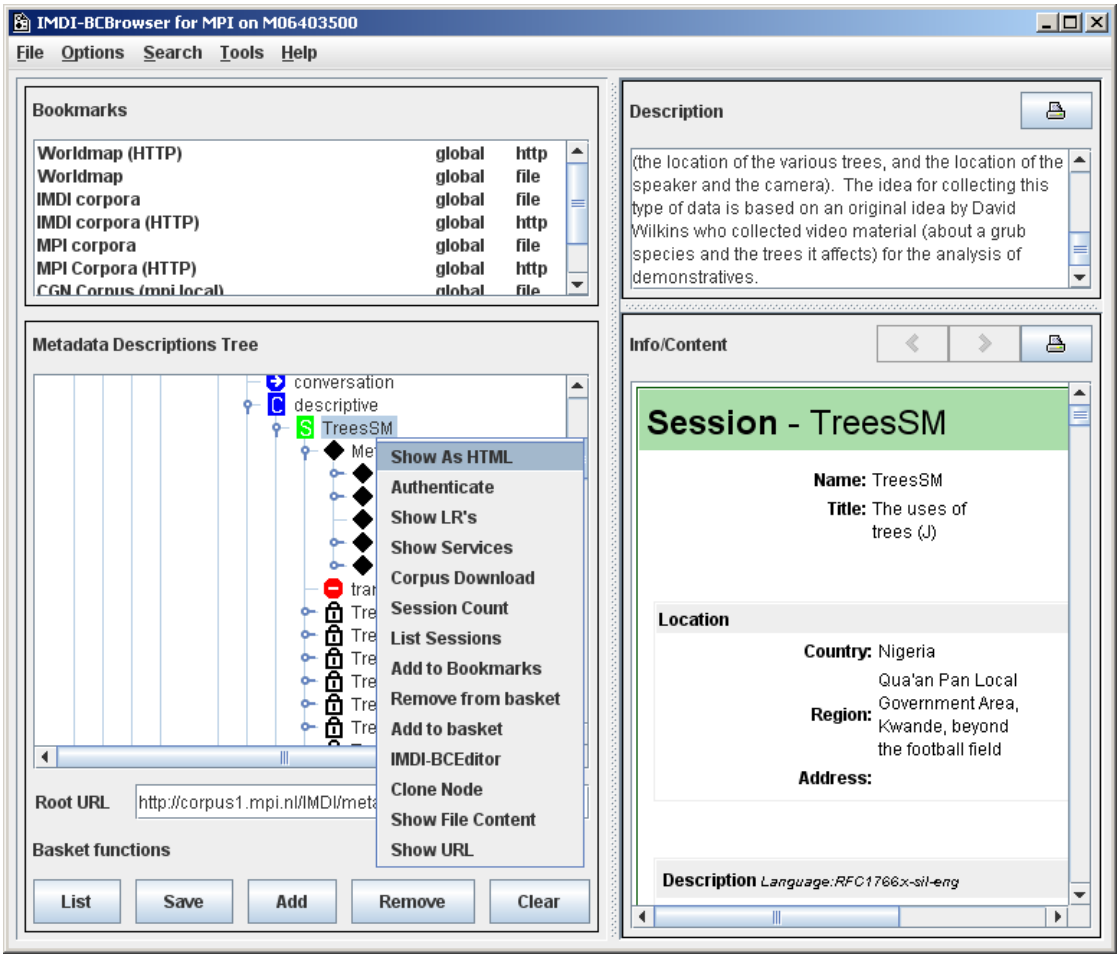


Figure 2.20. Show as HTML

- Choose IMDI-BCEditor from the menu. The IMDI Editor opens with the corresponding metadata file, e.g.:

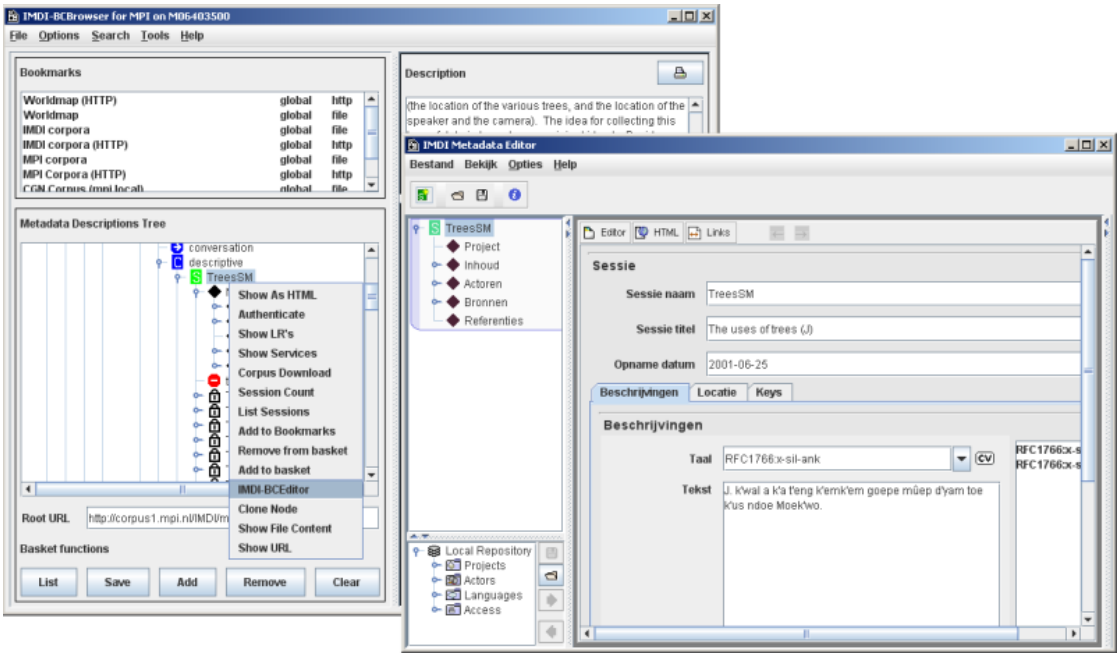


Figure 2.21. Open with IMDI-BCEditor

Note

You can update the metadata information in the IMDI Editor. However, in most archives, it is not possible to save the updated file directly to the server. Please save it to the local disk, and contact your corpus manager. In any case, please contact your corpus manager before you make any substantial changes to make sure that your changes do not interfere with the overall corpus management.

2.3.4.2. Digitized media files and transcription/annotation files

A session node also contains links to digitized media files (i.e., to audio and/or video files), and to transcription/annotation files. These files can be accessed either through the session node, or the media file node or the transcription/annotation file node.

The following options are available:

1. Right-click on a selected media file node to open the pull-down menu, then choose Windows Media Player, WMplayer, Audition or whatever program your operating system offers, from the menu. The media file will be opened in the appropriate, e.g.:

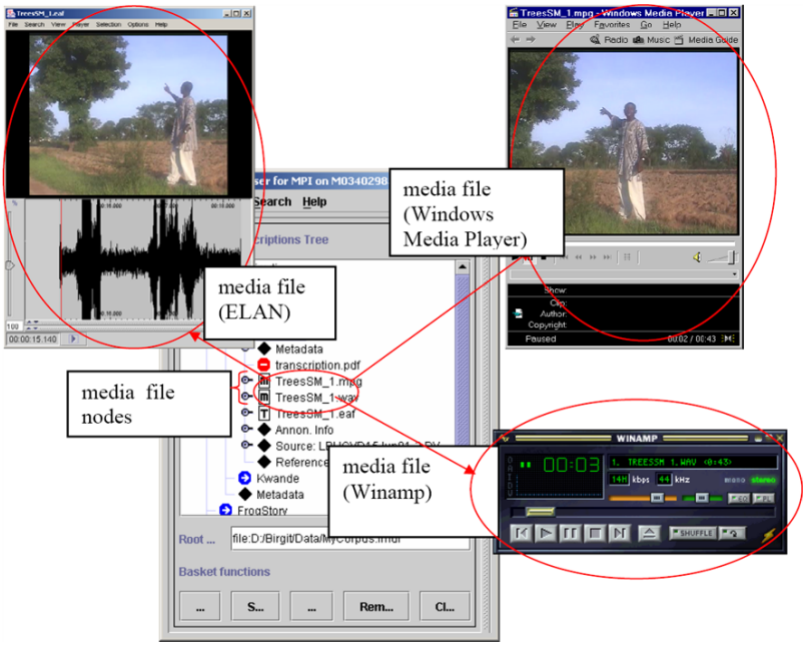


Figure 2.22. Opening media

In order to view digital images (`*.gif`, `*.jpg`) right-click on the corresponding node and select # Show #. The image will be displayed in the right window, titled Info/Content:

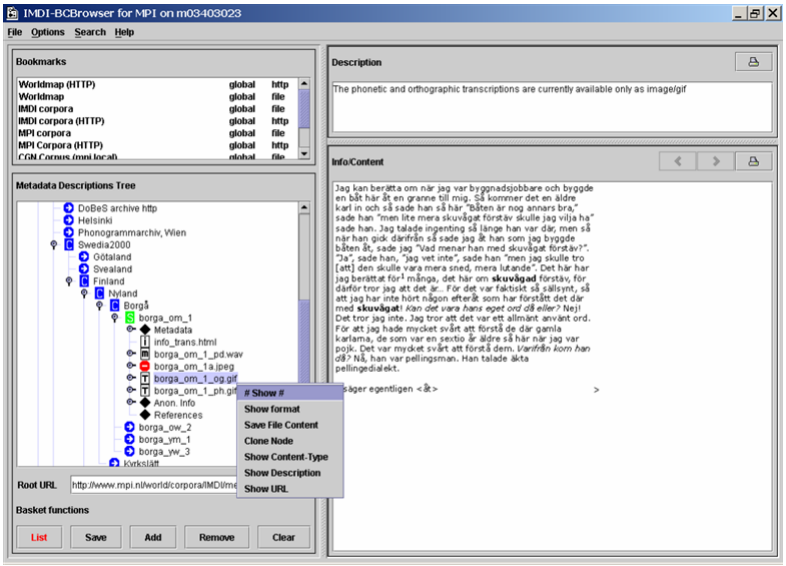


Figure 2.23. Showing images

2. Right-click on a selected session node to open the pull-down menu. If the annotation of that session was done in ELAN (i.e., if there is an `*.eaf` file), the pull-down menu will display the option ELAN. Select this option to view the media file together with its annotations.

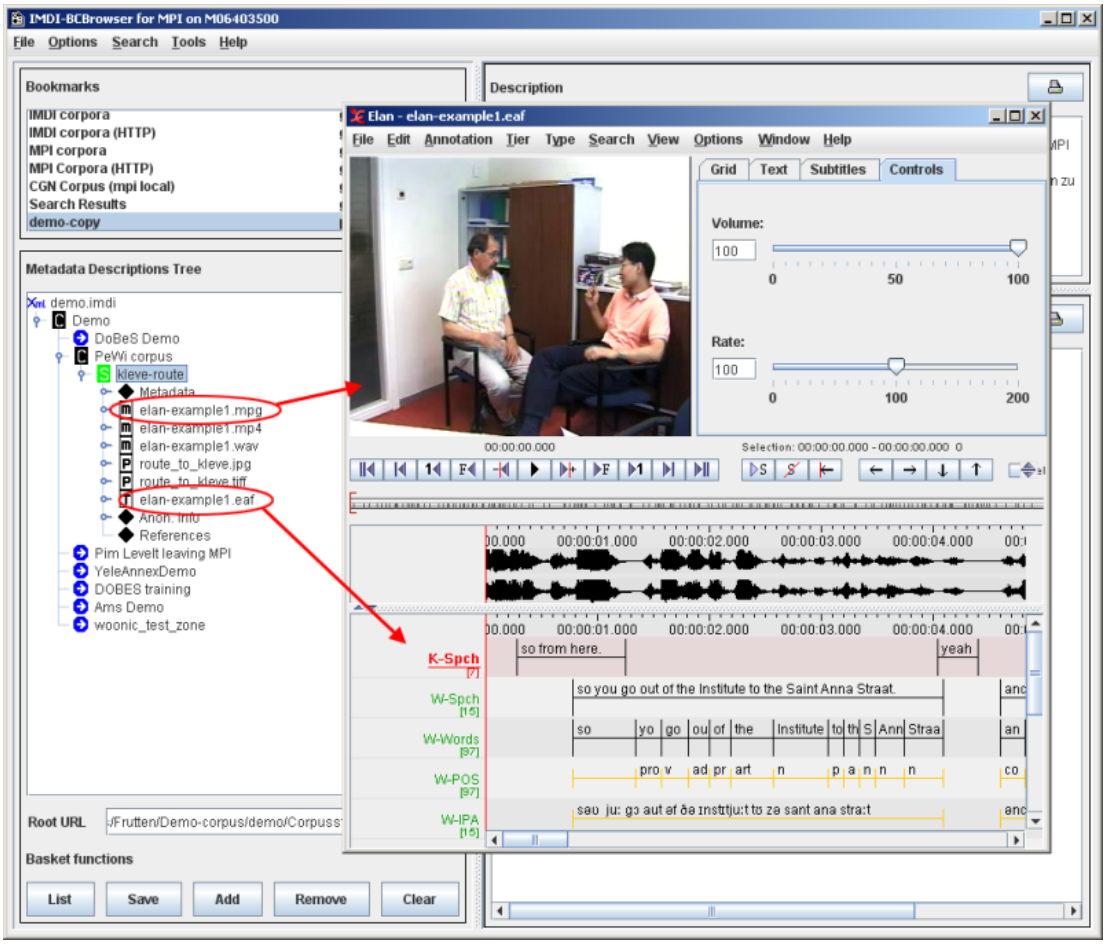


Figure 2.24. ELAN

3. Right-click on a selected annotation file node to open the pull-down menu, then choose Show from the menu. The file content is displayed in the Info/Content panel, e.g.:

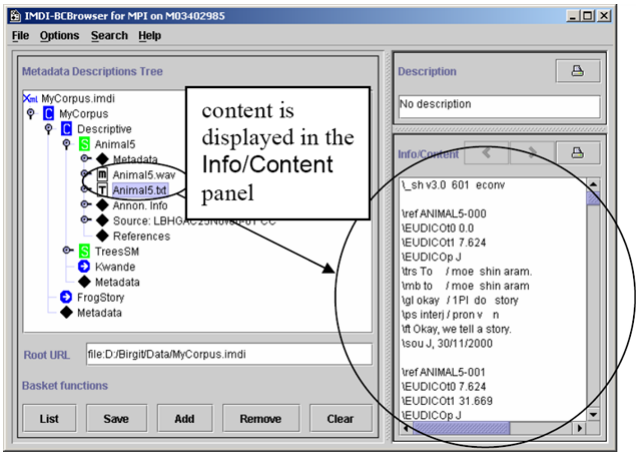


Figure 2.25. Show annotations

Note

There are no restrictions on the format of your annotation files (ELAN, Media Tagger, Shoebox, CHAT, etc.).

Note

You can constantly update your annotation files (provided you do not change their file names), which allows you to reflect changes in your analysis. But note that, for the moment, only ELAN can be directly accessed via the IMDI Browser. If you want to edit annotation files of a different format, please navigate to the physical location of that file (by using, e.g., Windows Explorer).

2.3.4.3. Info files

Both corpus and session nodes can contain links to info files. These files provide general background information about the corpus or the individual session.

If such a file is an HTML file, click on it: its content will automatically be displayed in the Info/Content panel. If the file is a PDF file, click on it and then right-click on it to open the pull-down menu. Choose Open with Acrobat from the pull-down menu. Acrobat Reader will open up to display the file.

For example:

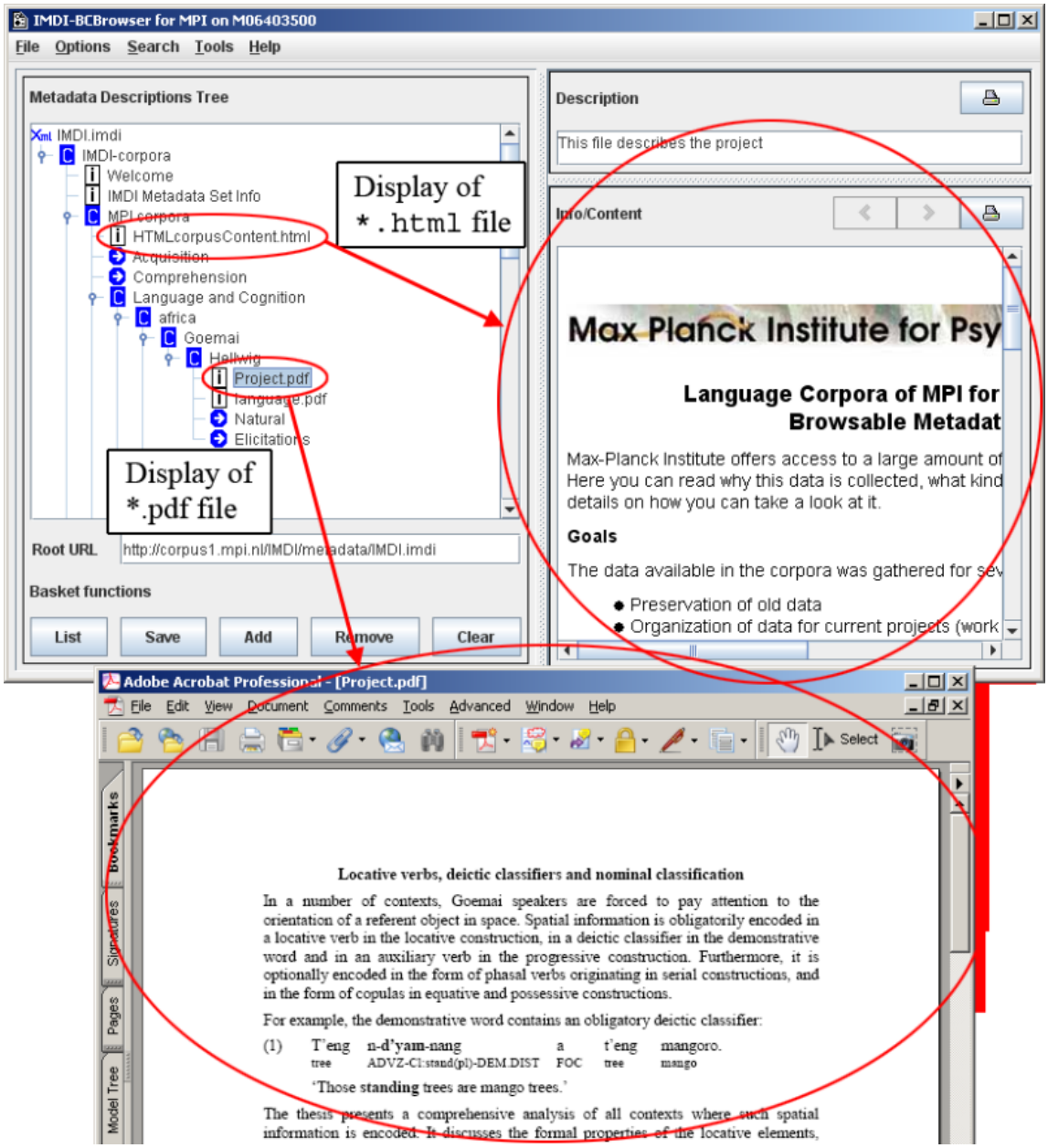


Figure 2.26. Info files

2.4. Info/Content panel

The Info/Content panel displays information about corpus, session and file nodes. To read the information, click on the node in the Metadata Descriptions Tree panel, e.g.:

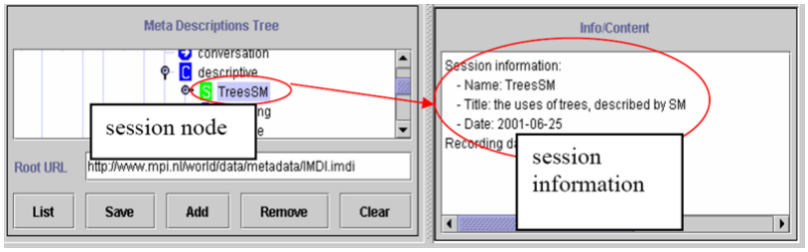


Figure 2.27. Info/Content panel

Note

The information is displayed in form of searchable keywords. For prose information, look at the Description panel (see Section 2.5).

The Info/Content panel can display additional information. To view this additional information, do the following:

1. In the Metadata Descriptions Tree panel, click on the node whose information you want to view. It will be highlighted in blue.
2. Right-click on the highlighted node.
3. Select one of the following options from the pull-down menu (see Section 2.3.1):
   - Bookmark Info: Displays information about the bookmark.
   - List Sessions: Lists all sessions contained under this node.
   - Session Count: Displays the number of sessions contained under this node.
   - Show: Displays the file content.
   - Show As HTML: Displays the file in an HTML format.
   - Show File Content: Displays the file in an XML format.
   - Show format: Displays the file format.
   - Show Info: Displays the file content.
   - Show LR's: Displays the directory information for all files contained under this node.
   - Show Services: Displays the name and directory information of the file needed for searching through the metadata.

   The corresponding information is displayed in the Info/Content panel.

The Info/Content panel also displays the world map (see Section 2.3.2). And it displays Backward and Forward buttons (that allow you to move to the previous or subsequent view) and a Print button (that allows you to print the information displayed in the panel).

2.5. Description panel

The Description panel displays a description of corpus, session and file nodes. To read the description, click on the node in the Metadata Descriptions Tree panel, e.g.:

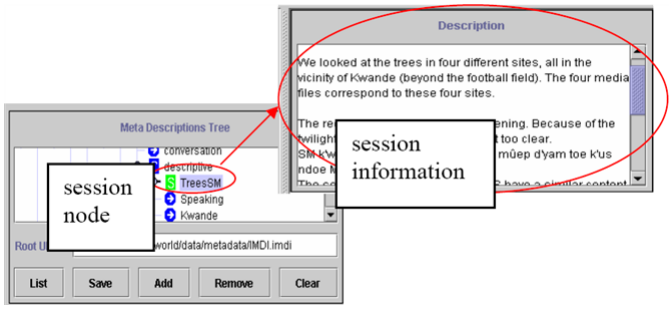


Figure 2.28. Description panel

Note

The Description panel displays prose information. For searchable keywords, look at the Info/Content panel (see Section 2.4).

The Description panel can display additional information. To view this information, do the following:

1. In the Metadata Descriptions Tree panel, click on the node whose information you want to view. It will be highlighted in blue.

2. Right-click on the highlighted node.

3. Select one of the following options from the pull-down menu (see Section 2.3.1).

   - Show Content-Type: Displays the file format.

   - Show Description: Displays the file description.

   - Show URL: Displays the directory information for the metadata file.

   The information is displayed in the Description panel.

The Description panel also displays a Print button (that allows you to print the information displayed in the panel).

2.6. Downloading a corpus

Copying a corpus - partially or a whole (sub)corpus - can be performed with the option Corpus Download. It is activated by right clicking on a corpus node and selecting Corpus Download from the context menu:

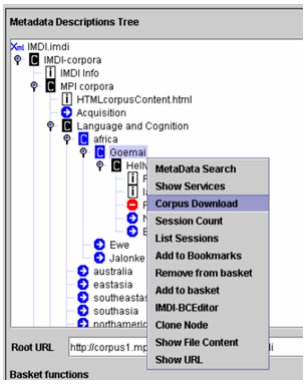


Figure 2.29. Corpus download

Doing so will open the Copy Dialog window:

Figure 2.30. Copy Corpus Dialog

After the files have been copied, you will find an exact clone of the selected subcorpus on your local machine, i.e. including the structure (directories).

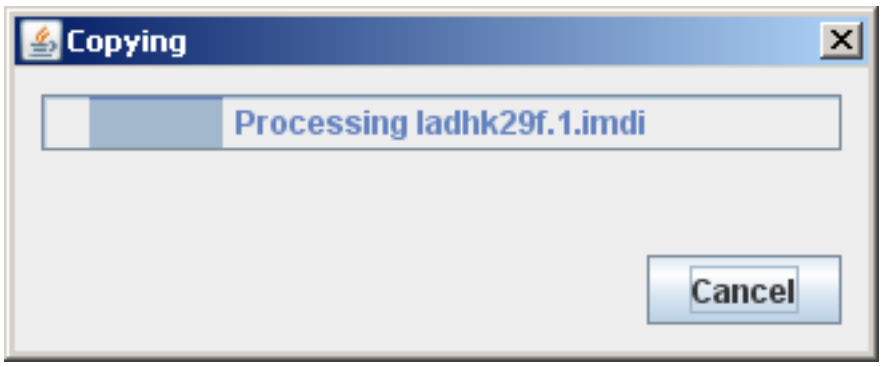


Figure 2.31. Copying

Note

Even if you don’t select any resource in the Proposal window, the directories and `*.imdi` files still are copied.

Chapter 3. Searching data (IMDI Metadata Search Tool)

To start the IMDI Metadata Search Tool, do one of the following:

1. Use the menu items (see Section 2.1.3).
    a. Click on Search.
    b. Click on Metadata Search.
2. Use the Metadata Descriptions Tree panel (see Section 2.3).
    a. Go to the Metadata Descriptions Tree panel.
    b. Double-click on a node to open it.
    c. Click on it to select it. It will be highlighted in blue.
    d. Right-click on it to open a pull-down menu.
    e. Select MetaData Search from the pull-down menu.

The following screenshot illustrates the Default search in the IMDI Metadata Search window:

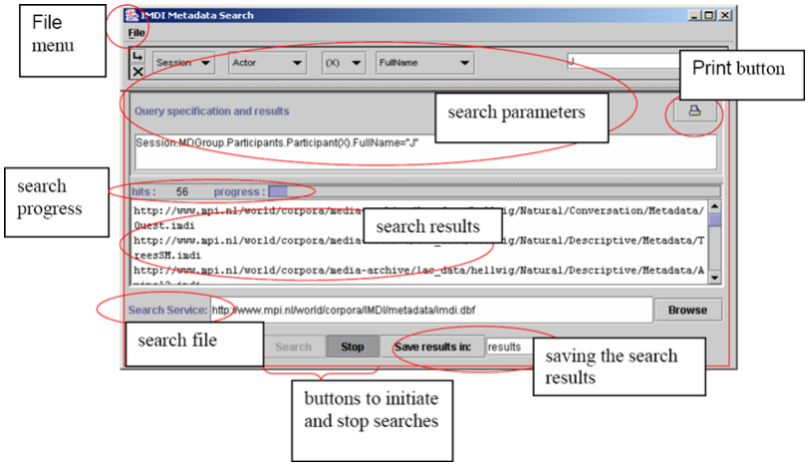


Figure 3.1. Metadata Search windows

Except the Default search (see Section 3.2.2), there is also a Key word search (described in Section 3.2.1) and a Advanced search (described in Section 3.2.3).

This section introduces the search options. The following options are available:

1. Specifying the corpus to be searched (Section 3.1).
2. Specifying the search parameters (Section 3.2).
3. Changing the profile (Section 3.3).
4. Displaying the search results (Section 3.4).
5. Saving the search results (Section 3.5).
6. Printing (Section 3.6).

7. Exporting a local search database (Section 3.7).

8. Closing the IMDI Metadata Search window (Section 3.8).

3.1. Specifying the corpus to be searched

The corpus to be searched is displayed in the box to the right of Data base:. If you right clicked a node and chose MetaData Search in the context menu, this node is also at top of the window (e.g. 'within corpus "IMDI-corpora"' in Figure 3.1). At the same time, you cannot change the search database by clicking the Browse button. Changing the search database is possible if you have clicked Search > Metadata Search.

For metadata search to work, you need to enter the name and location of the `*.data` file that contains the metadata information of the corresponding corpus (see Section 3.1.1). Many corpora (at least the corpora on the server) provide you with predefined `*.data` files, but you also have the possibility to create your own files for locally stored corpora (see Section 3.1.2).

Note

As of version 3.0 of the IMDI browser, the `*.data` format has changed. Make sure you create them again in order to avoid problems while searching.

3.1.1. Selecting a `*.data` file

To enter a `*.data` file in the box to the right of Data base:, do the following:

1. Either: Type in its name and location.

2. Or: Make use of the Browse button to the right of this box.

3. Or: Make use of the IMDI Browser interface. In this case, do the following:

   a. In the Metadata Descriptions Tree panel (see Section 2.3), double-click on the corpus node to open it. Then click on the node to select it. It will be highlighted in blue.

   b. Do one of the following:

      i. In the Search menu, click on Metadata Search (see Section 2.1.3).

      ii. Or in the Metadata Descriptions Tree panel (see Section 2.3), right-click on the selected node; then select MetaData Search from the pull-down menu.

   In the IMDI Metadata Search window, the `*.data` file associated with the selected corpus node is automatically entered into the box, e.g.:

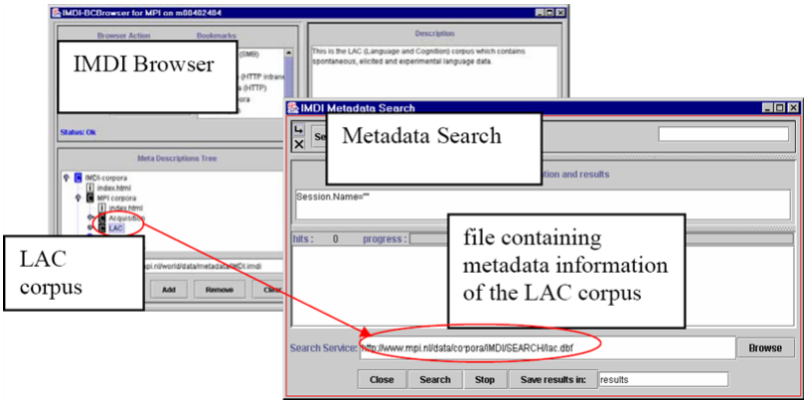


Figure 3.2. .data file association

Note

The name and location of a `*.data` file can be seen in the IMDI Browser. In the Metadata Descriptions Tree panel, double-click on a corpus node (to open it), click on it (to select it), and then click with the right-mouse button to open a pull-down menu. From the pull-down menu, select Show Services. The name and location of the corresponding `*.data` file is displayed in the Info/Content panel (see Section 2.3.1).

Note

Locally stored corpora can only be searched if there is a `*.data` file available. Such corpora are visualized in the Metadata Descriptions Tree panel by means of a white “C” on a black background. Corpora that have no corresponding `*.data` file are visualized by means of a white “C” on a blue background. If you click on a ‘non-searchable’ subcorpus that is part of a larger ‘searchable’ corpus, the whole corpus will be searched – not just the subcorpus. E.g.:

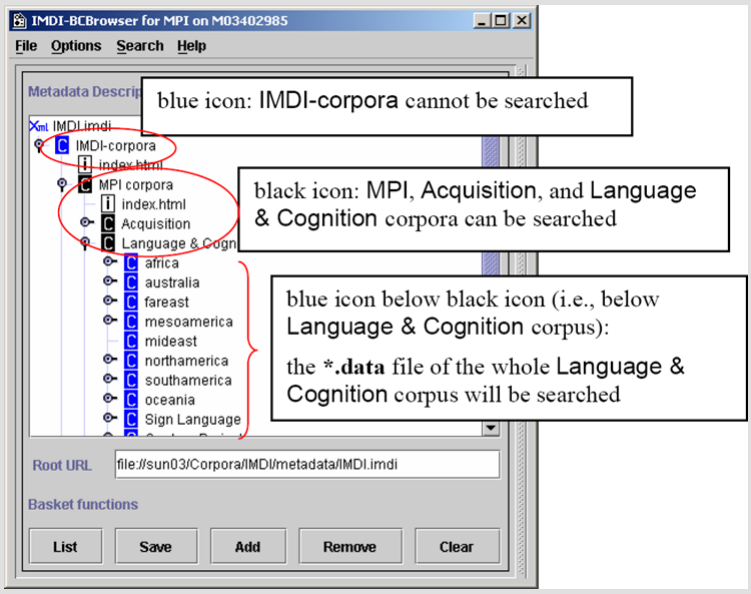


Figure 3.3. Searchable corpora

By default, metadata searches are conducted through an entire corpus. However, it is possible to limit your search to selected parts of a corpus. In this case, you have to manually select corpus and/or session nodes in the IMDI Browser. To select nodes, make use of the buttons Add, Remove and Clear at the bottom of the Metadata Descriptions Tree panel (see Section 2.3.3). Once you have selected node(s), metadata search is automatically restricted to the selected node(s). Again, you can only search parts of corpora that have a corresponding `*.data` file.

Note

If the search process does not yield the expected results, please check the following possibilities:

1. Make sure that the `*.data` file displayed next to Data base: is the correct file.
2. For searches through an entire corpus: make sure that you have not accidentally selected a part of the corpus. If you have selected a part, the List button at the bottom of the Metadata Descriptions Tree panel is highlighted in red color.

3. For searches through selected parts of a corpus: make sure that the list (displayed under the List button at the bottom of the Metadata Descriptions Tree panel) contains all relevant nodes.

3.1.2. Creating a new *.data file

Locally stored corpora that have a *.data file are visualized in the Metadata Descriptions Tree panel by means of a white "C" on a black background (see Section 3.1.1 above). However, if you want to search a corpus that does not have a *.data file, you must create your own. To do so, do the following:

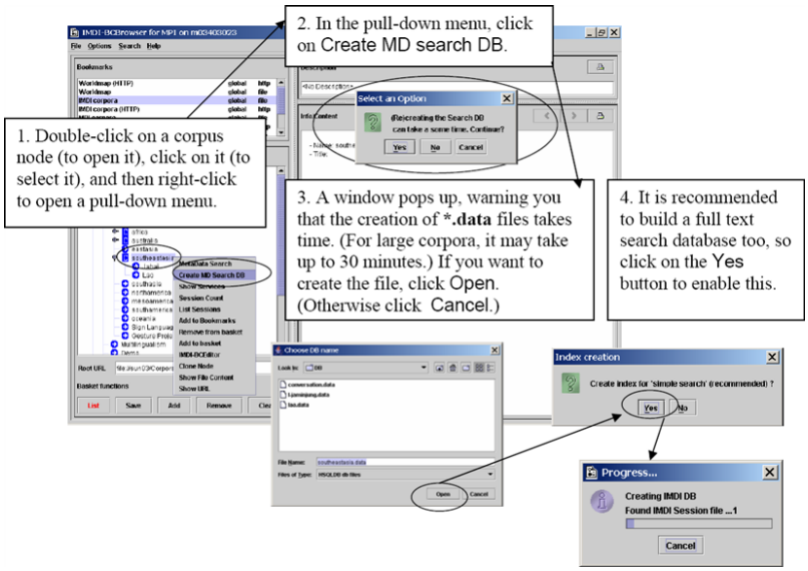


Figure 3.4. Create you own .data file

1. *Double-click on a corpus node (to open it), click on it (to select it), and then right-click to open a pull down-menu.*
2. *In the pull-down menu, click on Create MD search DB.*
3. *A window pops up, warning you that the creation of *.data files takes time. (For large corpora, it may take up to 30 minutes.) If you want to create the file, click Yes. (Otherwise click No.)*
4. *Enter a filename and click Save. (Or click Cancel to cancel.)*
5. *It is recommended to build a full text search database too, so click on the Yes button to enable this.*
5. *After the *.data file has been created, the corpus node changes its color from blue to black, and you can now search this corpus (see Section 3.1.1).*

Note

These *.data files are stored permanently in your local directory "user-directory \ IMDI-TOOLS \ DB".

3.1.3. Removing a *.data file

Once a local *.data file is created, it also can be deleted:

1. select Search > Metadata search
2. Choose Options > Local Imdi DB files from the dialog
3. Select the data-files that should be removed (you can select mutiple files by holding SHIFT or CTRL)
4. Press the Delete button. Now the `*.data` file is deleted.

3.2. Search types

There are three types of Metadata Search:

- Key word search (Section 3.2.1).
- Default search (Section 3.2.2).
- Advanced search (Section 3.2.3).

3.2.1. Key word search

As from version 3.0 the IMDI browser supports a Google-like search operation on a corpus. It allows you to perform a search on all fields of an imdi file. This can be done by clicking on the Key word search button in the search dialog window:

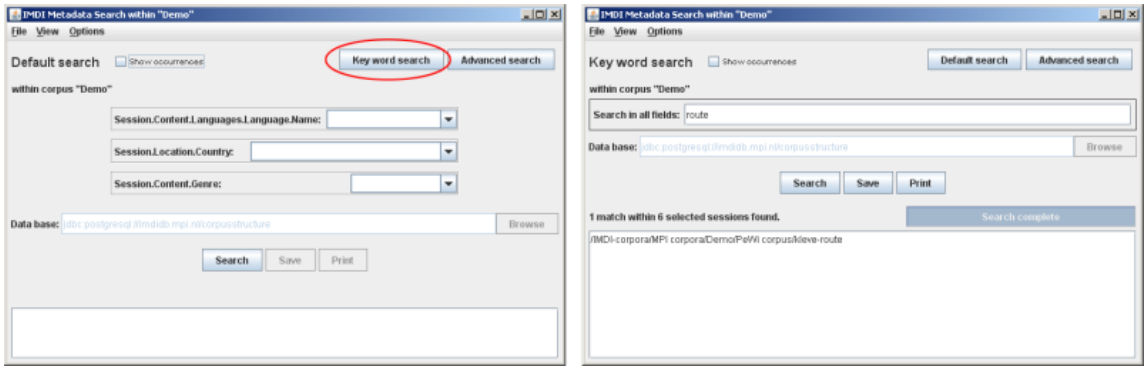


Figure 3.5. Key word search

Enter a search string net to Search in all fields. Click the Search button at the bottom of the IMDI Metadata Search window to start the search. During the search the number of “hits” (matches) to your search criterion is shown, as well as the progress towards completion of the search. Once the search process has started, you can use the Cancel button to stop the search.

Note

Because of the large amount of data that is searched, it may take some time before the search is completed.

3.2.2. Default search

When the Metadata Search window is opened, the Default search is displayed (see also Figure 3.1). Default search enables you to specify a query for three categories: Content Language Name, Country and Genre. For each category, the value to be searched can be selected from a pull-down menu.

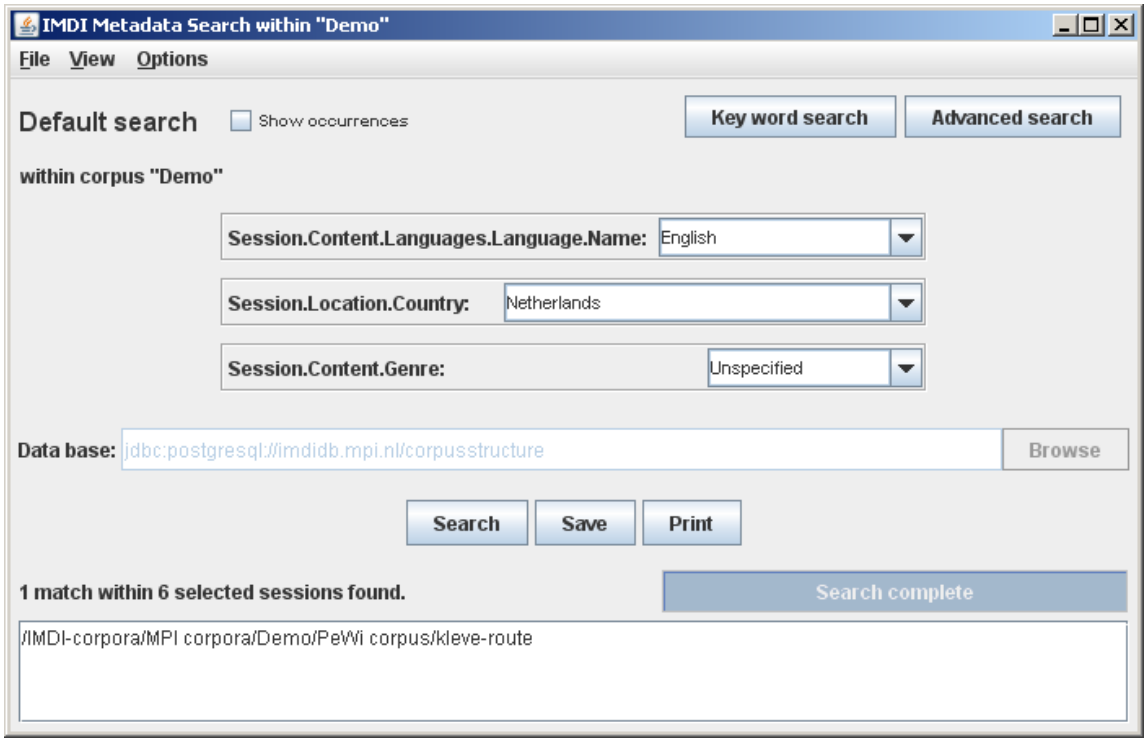


Figure 3.6. Default search

After selecting the values that are to be searched, click the Search button at the bottom of the IMDI Metadata Search window to start the search. During the search the number of “hits” (matches) to your search criterion is shown, as well as the progress towards completion of the search. Once the search process has started, you can use the Cancel button to stop the search.

Note

Because of the large amount of data that is searched, it may take some time before the search is completed.

3.2.3. Advanced search

In the Metadata Search window, click Advanced search to enter the advanced search dialog. In this dialog you can construct multiple contraint for your search.

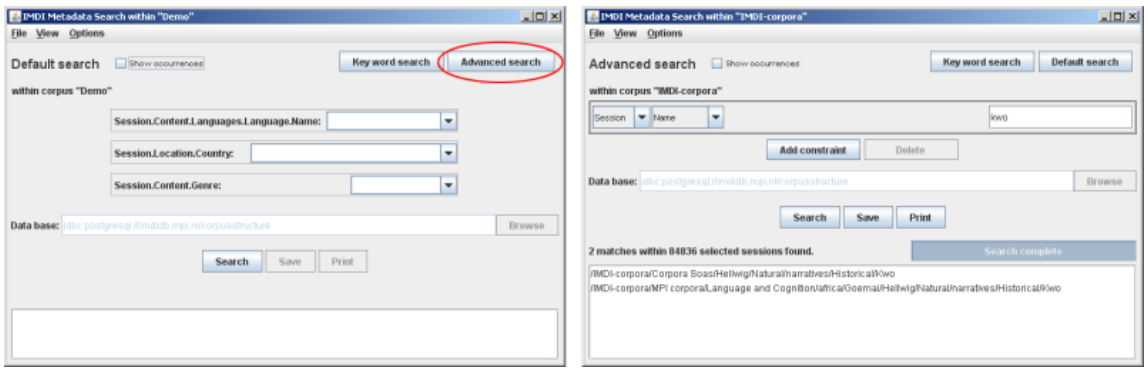


Figure 3.7. Advanced search

3.2.3.1. Selecting the category to be searched

The categories that can be searched are displayed in form of pull-down menus. By default, Session and Name are displayed. These two options allow you to search for all sessions with a particular name (see Figure 3.7).

Click on the pull-down menus to select other categories:

- The first category is always Session (i.e., you will always search for a session)
- The second category is chosen from the following list: Name, Title, Date, Location, Project, Keys, Content, Actor, Resource or References (i.e., you will search for a session that has a certain name, a certain title or a certain recording date, etc.).
- Depending on your choice for the second category, further pull-down menus will appear to narrow down the category.

Note

If too many categories and pull-down menus are added, the IMDI Metadata Search window cannot display them all. To increase its size, click on the full screen icon (in the top right corner of the window):

Some examples of metadata searches are given below:

1. The screenshot below illustrates a search for all sessions that were recorded in the Netherlands.

   In this case, Session and Location were chosen as the first two categories. A third pull-down menu appeared subdividing the category Location into Continent, Country, Region and Address. When Country was chosen, a fourth pull-down menu appeared displaying the names of all countries.

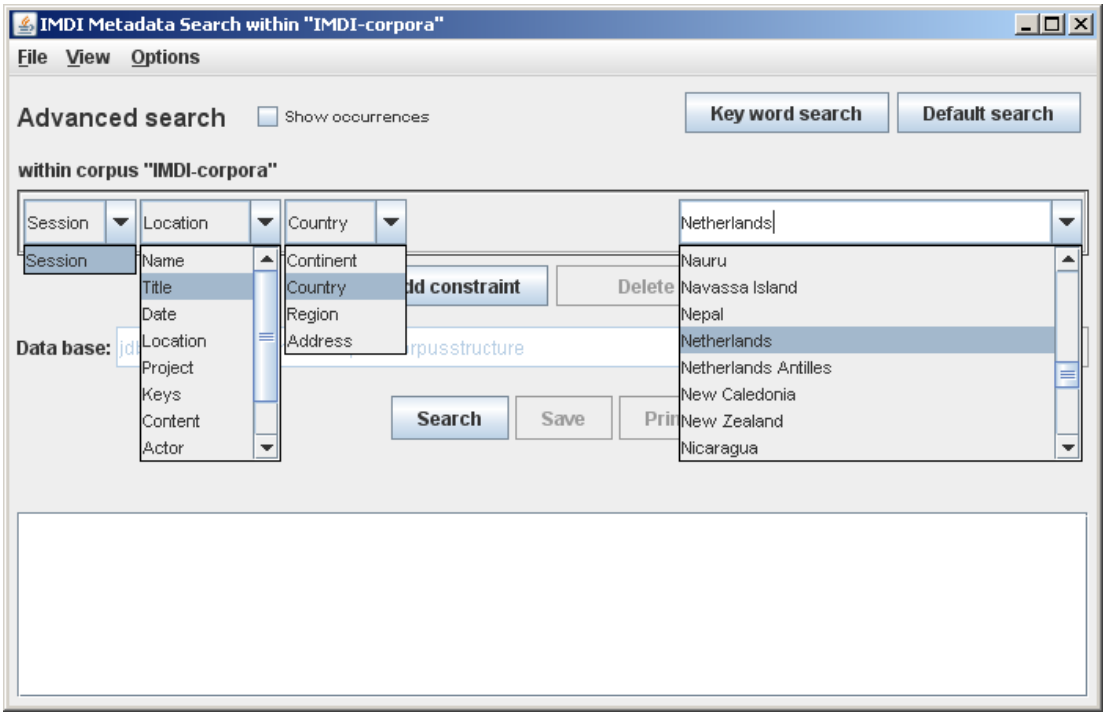


Figure 3.8. Query specification

2. Figure 3.9 illustrates a search for all sessions that were recorded after January 1, 2000 and before January 1, 2001.

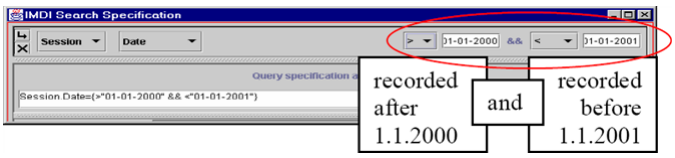


Figure 3.9. Date specification

In this case, Date was chosen as the second category (i.e., the date of the recording). Further pull-down menus appeared that allowed to specify the exact time period. The first pull-down menu has the following options:

- < (less than): Search for a date *before* the specified date;
- = (equal to): Search for the exact date;
- > (greater than): Search for a date *after* the specified date;

3. Find all sessions where gestures belong to the modalities that are described. So first select Session, then Content, Modalities and finally gestures.

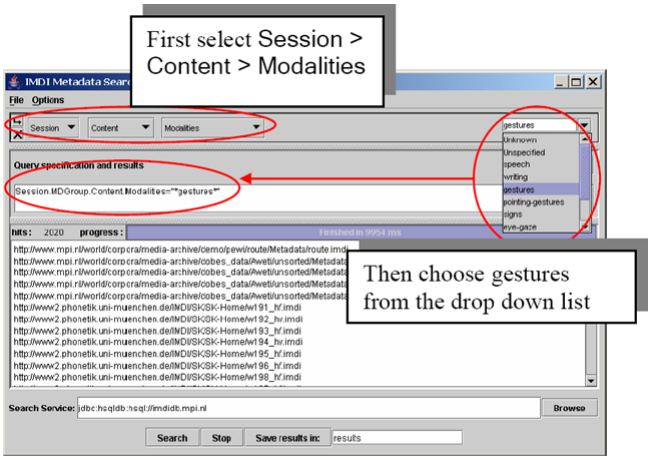


Figure 3.10. Select gestures as modality

3.2.3.2. Entering the search item

After you have selected the search categories, enter the search item into the box in the top right corner, and then press the key ENTER, e.g.:

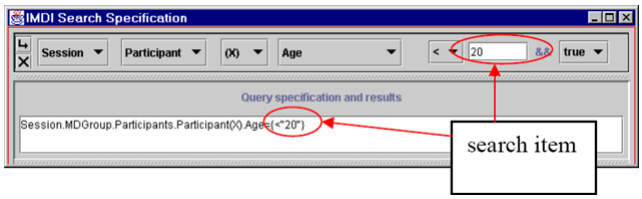


Figure 3.11. Entering the search item

Note

Some search items have to be entered manually, others can be chosen from a pull-down menu (see Section 3.2.3.1). If you enter the search item manually, you always need to press the key ENTER to commit the item.

3.2.3.3. Adding or deleting a search query

You can add and delete constraints as follows:

- Click Add constraint to add a constraint;
- Click Delete to delete the last constraint;

When two (or more) constraints are specified, the search process will find all sessions that satisfy constraint 1 *and* constraint 2 (*and* all subsequent constraints). For example, in Figure 3.12, all sessions will be found that contain utterances from (1) a male actor X *and* (2) a female actor Y.

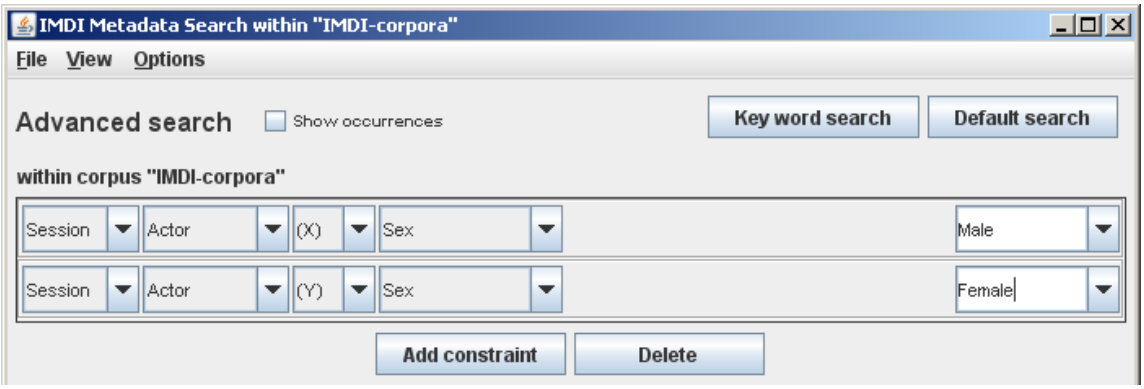


Figure 3.12. Additional constraint

Note

Variables such as X and Y are only relevant if you specify more than one constraint. In this case, they allow you to specify that the actor of constraint 1 (e.g., X) should either be different from the actor of constraint 2 (e.g., Y) (as in the illustration above), or that (s)he should be identical to the actor of constraint 2 (e.g., X). Please ignore these variables if you specify only one constraint.

3.2.3.4. Initiating and stopping the search

After entering one or more constraints, click the Search button at the bottom of the IMDI Metadata Search window to start the search. During the search the number of "hits" (matches) to your search criterion is shown, as well as the progress towards completion of the search. Once the search process has started, you can use the Cancel button to stop the search.

Note

Because of the large amount of data that is searched, it may take some time before the search is completed.

3.3. Changing the profile

Section 3.2.3.1 above has shown that the categories to be searched are displayed in form of pull-down menus. These categories are specified in a so-called "profile" (stored in the folder "user-directory \ IMDI-TOOLS \ Profiles \ local"). To change the profile, do the following:

1. Click on File menu.
2. Click on Change Profile. The Open dialog window appears and displays all available profiles, e.g.:

Figure 3.13. Changing the profile

3. Double-click on the profile you want to use. The IMDI Metadata Search window will display the chosen profile. E.g., the following illustration displays some categories that are specific to sign language research and that are defined in the profile "SignLanguage.Profile.xml":

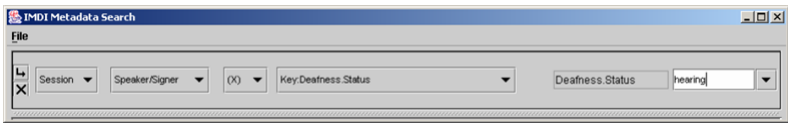


Figure 3.14. Categories of the new profile

Note

Note: For the moment, profiles have to be programmed by hand. In future versions of the IMDI tools, it will be possible to create profiles with the help of the IMDI (Metadata) Editor.

3.4. Displaying the search results

The box at the lower part of the window displays the results, i.e., the name and location of each session. If you selected Actor in the search constraint, the results will also include the Code(s) of the Actor(s) that fit the constraint. It helps you identify the actor(s) you are looking for.

You can access any search result by double-clicking it. This will open up an IMDI Browser window containing the selected session, its metadata descriptions and its media and annotation files.

3.5. Saving the search results

You can save the search results in the form of a corpus node. This corpus node is displayed in the Metadata Descriptions Tree panel of the IMDI Browser window. Do the following:

1. In the IMDI Metadata Search window, click on Save;

2. In the next dialog, select a directory and enter Click the button Save results in:. The results are saved under the specified name.

You can access the results through the IMDI Browser interface. Do the following:

1. In the IMDI Browser window, go to the Bookmarks panel (Section 2.2).

2. Double-click on Search Results.

The Metadata Descriptions Tree panel (see Section 2.3) displays the new corpus node below the node labeled Results.

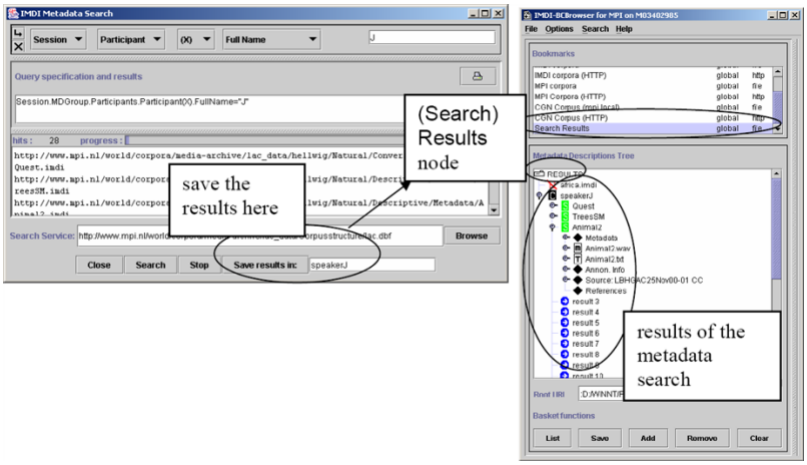


Figure 3.15. Accessing the search results

The new corpus node is treated like any other corpus node, i.e., it contains sessions, and the sessions contain metadata, media and annotation files.

You can remove the saved corpus node again. Do the following:

1. In the Metadata Descriptions Tree panel (see Section 2.3), click on the node to select it. It will be highlighted in blue.
2. Right-click on the selected node to open a pull-down menu.
3. Select Remove from the pull-down menu.

   The node will be removed.

Note

After selecting Remove, the icon of the removed node remains visible although it cannot be accessed anymore. The icon will be removed after exiting the node Results.

3.6. Printing

You can print the search parameters (see Section 3.1) together with the search results (see Section 3.4). To do so, click on the Print button.

3.7. Exporting a local search database

Suppose you have a collection of IMDI files from which you want to extract certain fields (e.g. all Actors) and export them in a tabular format. To achieve this, first make sure you have a local search database. If not, create one as described in Section 3.1.2. Then, in the Metadata Search dialog, choose Options > Local Imdi DB files. This will result in a screen as below:

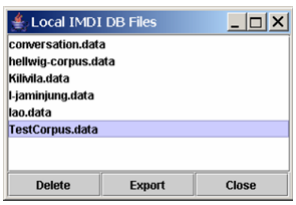


Figure 3.16. Local IMDI DB files

Select a local corpus node and click on Export. A save dialog will be displayed. Choose a directory to store the exported files.

In the end, about 12 files with the prefix IMDIMD- will be saved. Each one contains the contents of a specific subset of an Imdi file (e.g. all actors are stored in IMDIMD_ACTOR). The format used is CSV (tab-separated) which means you can easily import them into Excel or any other application. Advanced users might be interested in the fact that this format allows the import of the data in a relational database.

3.8. Closing the IMDI Metadata Search window

Click on File menu, then click on Close to exit the window.

Chapter 4. Changing files

When you want to make changes to a large amount of IMDI files, editing one after the other might be a bit awkward. Therefore the IMDI tools provide a system that supports the automatic modification of groups of imdi files. This works as follows:

- The user creates a so called update stylesheet, using the IMDI editor. This file contains a description of the desired changes. (See section 1.2.3.4 in the IMDI editor manual, *Create Update StyleSheet*)
- With the IMDI browser you can apply the changes as described in the stylesheet to a set of files.

Note

If you want to change a large amount of files, be sure to test your stylesheet on a smaller set in advance. Otherwise a small mistake might corrupt many files.

A stylesheet made with the IMDI editor can be activated by following these steps:

- Select the Options > Enable BatchModifier menu and make sure this option is enabled.
- Make a selection:
  - Select a single IMDI file or a corpus node through File > Open file.

Note

If you select a corpus node, all child nodes will be affected!

Note

A corpus node can only be used if there exists an associated search database (i.e., the C icon should have a black background). To create a search database, see Section 3.1.2)

  - Or type in a directory path in the Root URL text field and press enter, e.g. d:\tmp. In this case the stylesheet will be applied to all imdi files found in this directory.

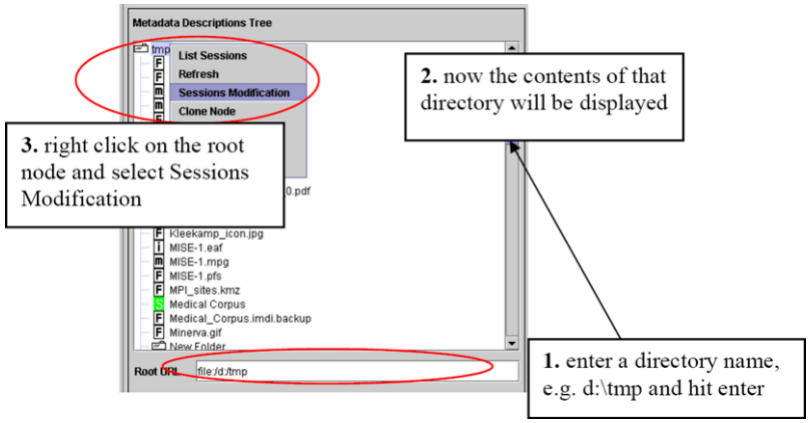


Figure 4.1. Entering a root url

1. *Enter a directory name and press* Enter.
2. *Now the contents of that directory will be displayed.*
3. *Right click on the root node and select Sessions Modification*

- Right click on it and select Sessions Modification
- Select the stylesheet you want to use and click on Apply
- Now the desired changes will be made to the selected file(s).

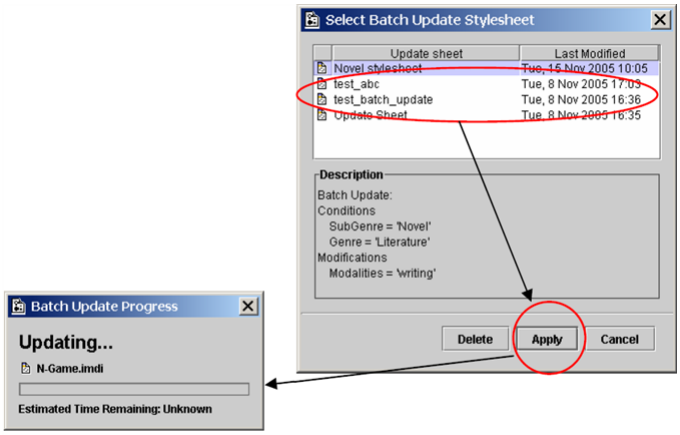


Figure 4.2. Select batch update stylesheet

Note that for all of the files to which the update stylesheet is applied a backup file (without the alterations) is created with the extension .old.1 (and .old.2 after another update, etc.).

Appendix A. How to setup an IMDI metadata Database server

The IMDI metadata DB server is based on the HSQLDB[1] software. The code for running the daemon is contained in "hsqldb.jar" and it is started from a script file "startIMDIdb.sh" . In this startup script information is specified where to find the java installation and the DB files. The DB server uses a number of DB files that hold the RDB tables with metadata. One of these files also contains configuration data that can be edited. The configuration information consists a.o. of the different corpus root url's that should be changed when the corpus changes (physical) location.

The DB files themselves are created from the IMDI browser just as the special DB files for local use are created (see the manual). With this difference that first some configuration information has to be specified. This is done by selecting:"Options" - "Preferences" - "DB creation". The resulting panel offers the following options:o

1. SeparateKeyValueTables: set this to false if the DB files concern a single corpus that systematically uses the same set of user defined keys. Default is true.

2. Include Descriptions: if true the information in the "free text" description elements is included in the DB. Default is false.

3. Corpus Root: The part of the URL path seen by the IMDI brower that all IMDI files of the corpora have in common. This information is used to make the path names in the DB files relative.

4. Remote Corpus Root: The URL path seen by a www-browser that all IMDI files of the corpus have inc ommon.

5. UNIX Local Corpus Root: The URL path seen by a (UNIX) NFS client that all IMDI files of the corpus have in common.

6. Windows Local Corpus Root: The URL path seen by a (Win) SMB client that all IMDI files of the corpus have in common.

This information concerning rool paths is needed if the DB files are to be used by both remote as well as local clients.

Once the DB files have been created, the DB daemon can be started with the startup script., With the currently use version (1.7.1) of the HSQL DB, the daemon can only handle one DB.

Finally the corpus topnode should be made to point to the DB daemon by modifying the topnode "SearchService" and "CorpusStructureService" attributes. These should both have the value: jdbc:hsqldb:hsql://hostname/ where hostname is the name of the host running the HSQL DB server. (N.B. This reference format will change with the next version of the HSQLDB software)

The command file for starting the IMDI DB server on Linux

```
#!/bin/sh
#This is the startup script for the HSQLDB IMDI DB server
#modify DBDIR to the directory where the DB files can be found
#modify JAVA_HOME to specify the java installation dir
#The hsqldb jar should be in the DBDIR directory
#
#
DBDIR=/data/corpora/IMDI/DB
JAVA_HOME=/usr/lib/java/jre
```

See http://hsqldb.sourceforge.net/ [???]

```
cd $DBDIR
$JAVA_HOME/bin/java -mx200M -classpath hsqldb__V1.7.1.jar org.hsqldb.Server \
 -database imdi 1>&2 > HSQLDB.log
```