

implementingDITA™

Notes from a real-world DITA Open Toolkit project



Anna van Raaphorst
Richard H. (Dick) Johnson
Authors of the *DITA Open Toolkit User Guide and Reference*

VR Communications, Inc.
www.vrcommunications.com

August 2006

Contents

- DITA and DITA Open Toolkit Overview
 - What is DITA?
 - What is DITA Open Toolkit?
 - Why use DITA?
 - Is DITA strategic?
- Overview of the *DITA Open Toolkit User Guide and Reference* project (“You can do this, too!”)
- Demo: Working environment, source and output files
- Evaluating DITA and the Toolkit for use in your environment
- Getting started
- Demo: Processing (building) DITA source files to multiple output environments
- Troubleshooting the build process
- Demo: Using the free debugging and reporting tools
- Production notes on other topics
- For more information

DITA and DITA Open Toolkit overview

What is DITA?

- XML-based, end-to-end architecture for authoring, producing and delivering information (often called *content*)
- Information is typically:
 - technical or scientific in nature
 - published as online help, product support portals or print-ready PDF files
- OASIS standard

Why use DITA?

- Publish to a wide variety of environments and platforms (DocBook, Eclipse content, Eclipse help, HTML Help (chm), JavaHelp, PDF, troff, Word RTF, XHTML)
- Share and reuse information (collaborate)
- Reduce writing, publishing and translation costs
- Produce more accurate and usable information
- Easy to specialize (if you really need to)

What is DITA Open Toolkit?

- Implementation of OASIS Technical Committee's specification for DITA DTDs and schemas
- Transforms DITA content (maps and topics) into deliverable formats
- Can be used in Windows, UNIX/Linux, MAC OS environments (anywhere that Java can run)

Why is DITA strategic?

- All the advantages of XML
- Promotes minimalist, topic-based, structured communication solutions
- Multi-channel, multi-language
- Easy to specialize (if you really need to)
- Cheap and relatively straightforward to get started
- Vendor interest and activity will lead to more sophisticated solutions

Project overview

What is the project?

- *DITA Open Toolkit User Guide and Reference* (release 1.2.2)
- Volunteer contribution to OASIS and the DITA community
- Bookmap specialization, 20 chapters, 300 topics
- Released August 10, 2006
- Will be updated to Toolkit 1.3 level and re-released in October
- DITA community input and comments welcome

Document contents

- Getting started
- Getting information
- Evaluating DITA and DITA Open Toolkit
- Installing, upgrading
- Setting up your working environment
- Processing and publishing to all 9 target environments
- Troubleshooting the build process
- Creating topics and maps, sample files
- Linking, customizing, distributing, reusing, migrating
- DITA core vocabulary (controlled vocabulary)

Why we did the project

- Significant interest and prior involvement with structured writing, content management, scripting, and programming
- Way for us to gain knowledge and experience quickly
- Help the DITA community
- Opportunity to use our individual skill sets in a collaborative effort



Key challenges

- Learning and writing simultaneously – for example:
 - How to choose authoring tool
 - Unfamiliar build environment
 - Some gains, some losses in functional richness
- How to collaborate effectively
- Change of initial project focus



Demo

Working environment,
source and output files

Evaluating, getting started

Evaluating, getting started

1. Be honest (can you answer “yes” to these questions?)
 - Do you have strong motivation and a champion with money?
 - Do you have the skills required (architecture, communication, technical)?
 - Can you collaborate effectively?
 - What are your key target outputs and will you be satisfied with the results (or will you need to customize)?
 - Do you have appropriate demo, prototype, and end-to-end projects in mind?
 - Do you have the time and patience to be successful?
2. Plan and architect your demo and prototype projects
 - Talk to others who have done similar projects
 - Get multiple people involved, including an information architect
3. Choose a (temporary) authoring tool
 - Start with with a free or inexpensive “DITA-aware” editor
 - Set it up to use the DTD catalog (if necessary)

Getting started, continued

4. Install the Toolkit and set up your working environment
 - Check the system requirements carefully (components must be compatible)
 - Separate your source and output files from the Toolkit files
5. Do a demo project
 - Single map, 12-20 files (include all information types), individual or two-person project
 - Add tables, definition lists, metadata, filtering, indexing, simple linking
 - Publish to all your key target environments
6. Do a prototype project with real content
 - Multiple maps, 20-50 files, group project
 - Publish to all key target environments
 - Add complex linking in stages

Demo

Processing (building) files

Debugging and reporting tools

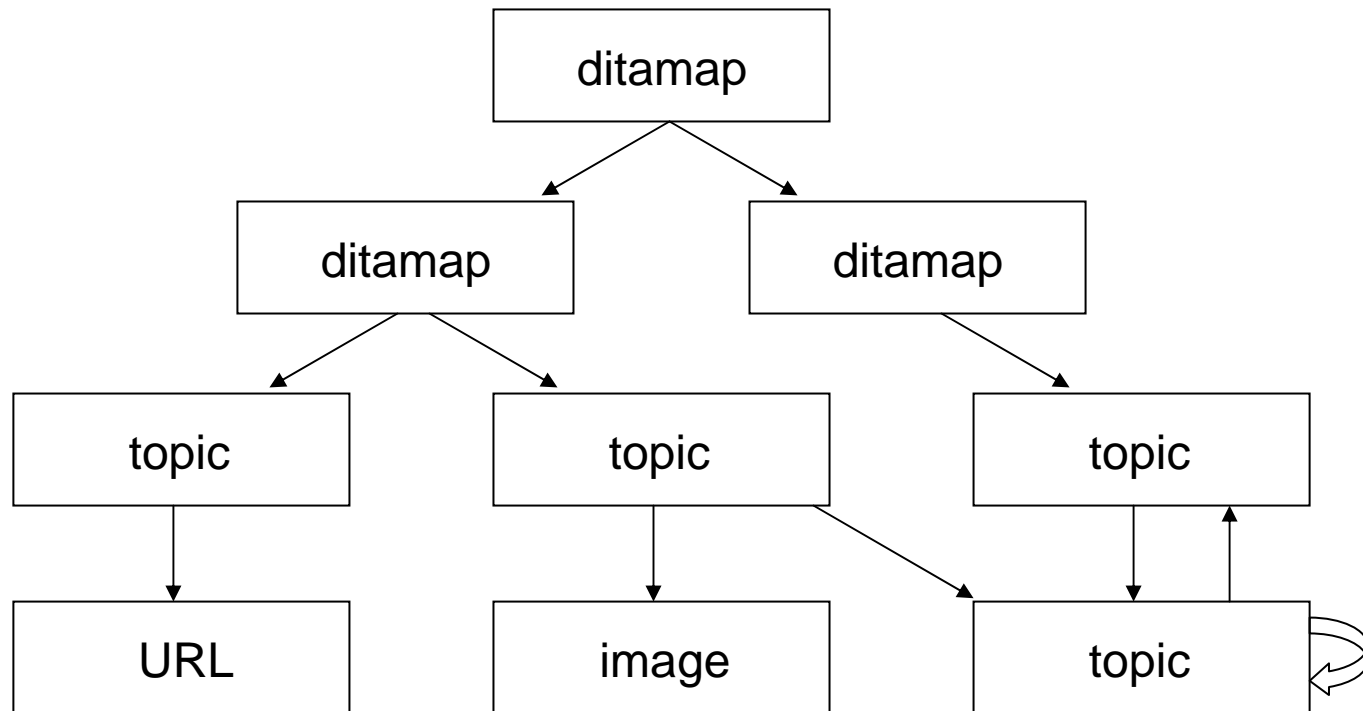
Why create DITA tools?

- DITA projects are made up of many related files defined by a DITA map
- Frequently need to perform operations on **all** files in a DITA map, for example:
 - Scan all files for file/URL reference errors
 - List files, directories and references in map
 - List usage of IDs and metadata keywords
 - List file author/copyright information
 - Perform global string list/replace
 - List all open draft questions
- **Cost of vendor authoring/CMS tools**

How to develop DITA tools?

- What makes tool development possible:
 - DITA files are well-formed and valid XML
 - By parsing the XML, tools can “walk” the file tree
 - Can use XPath queries for selecting individual file contents
- Our tools (available free with the *DITA Open Toolkit User Guide*) were written in PHP using its XML and file processing capabilities
- PHP is free, open source, runs on all platforms

DITA project file tree: the key to debugging and reporting



Our DITA Tool Set

- ditadebug – check for bad references
- ditalinks – verify all URLs referenced
- ditaauthors – list topic authors and copyright holders
- ditaids – list IDs used by files
- ditakeys – list metadata keywords used in files
- ditaedit – global string find/replace

ditadebug

- Checks for reference errors
- Lists:
 - Directories in map
 - Files in map
 - Files not used
 - References in map
- Output suitable for import into spreadsheet

Cryptic, misleading error messages

Toolkit error message:

```
[DOTJ013E][ERROR] Failed to parse the referenced file  
'installing\indows_settingenvvariables.dita'
```

Output from the debugging tool:

```
Error, file  
C:\DITAOT_UGRef_SOURCE\installing\indows_settingenvvariables.dita  
Does not exist! Bad reference:  
C:\DITAOT_UGRef_SOURCE\installing\windows_installingant.dita  
-> indows_settingenvvariables.diat
```

Name folding errors

No debugging help from the Toolkit (but some serious output errors)

Output from the debugging tool:

Name folding error, file

C:\MY_DITA_SOURCE\samples\garage\tasks\../Concepts/oil.dita

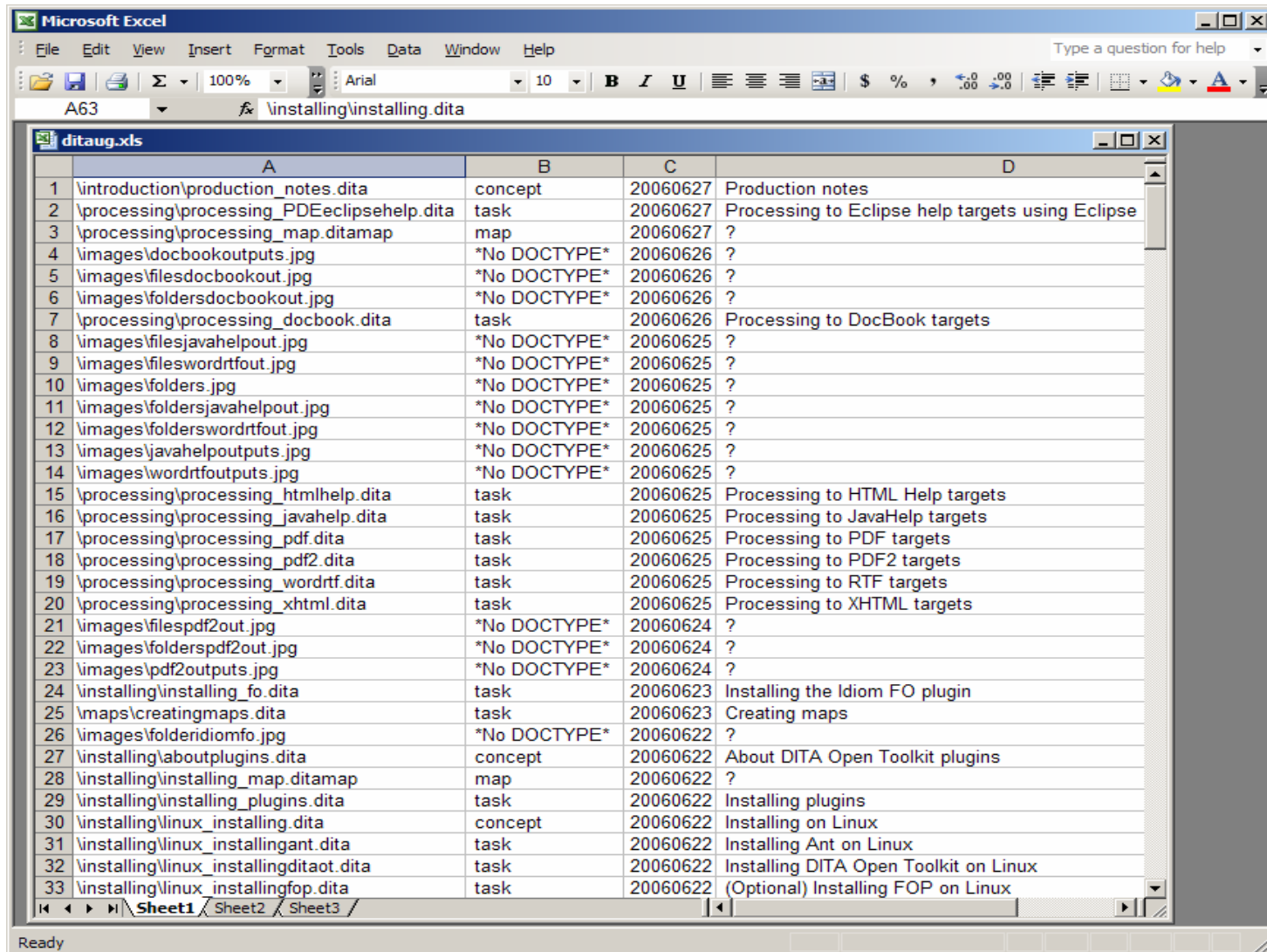
Does not exist!

Bad reference:

C:\MY_DITA_SOURCE\samples\garage\tasks\changingtheoil.dita

->../Concepts/oil.dita

File list: list of changed files



The screenshot shows a Microsoft Excel window titled "ditaug.xls" with a spreadsheet containing a list of files and their associated tasks. The spreadsheet has four columns labeled A, B, C, and D. The data is as follows:

	A	B	C	D
1	\\introductio\production_notes.dita	concept	20060627	Production notes
2	\\processing\processing_PDEclipsehelp.dita	task	20060627	Processing to Eclipse help targets using Eclipse
3	\\processing\processing_map.ditamap	map	20060627	?
4	\\images\docbookoutputs.jpg	*No DOCTYPE*	20060626	?
5	\\images\filesdocbookout.jpg	*No DOCTYPE*	20060626	?
6	\\images\foldersdocbookout.jpg	*No DOCTYPE*	20060626	?
7	\\processing\processing_docbook.dita	task	20060626	Processing to DocBook targets
8	\\images\filesjavahelpout.jpg	*No DOCTYPE*	20060625	?
9	\\images\fileswordrtfout.jpg	*No DOCTYPE*	20060625	?
10	\\images\folders.jpg	*No DOCTYPE*	20060625	?
11	\\images\foldersjavahelpout.jpg	*No DOCTYPE*	20060625	?
12	\\images\folderswordrtfout.jpg	*No DOCTYPE*	20060625	?
13	\\images\javahelpoutputs.jpg	*No DOCTYPE*	20060625	?
14	\\images\wordrtfoutputs.jpg	*No DOCTYPE*	20060625	?
15	\\processing\processing_htmlhelp.dita	task	20060625	Processing to HTML Help targets
16	\\processing\processing_javahelp.dita	task	20060625	Processing to JavaHelp targets
17	\\processing\processing_pdf.dita	task	20060625	Processing to PDF targets
18	\\processing\processing_pdf2.dita	task	20060625	Processing to PDF2 targets
19	\\processing\processing_wordrtf.dita	task	20060625	Processing to RTF targets
20	\\processing\processing_xhtml.dita	task	20060625	Processing to XHTML targets
21	\\images\filespdf2out.jpg	*No DOCTYPE*	20060624	?
22	\\images\folderspdf2out.jpg	*No DOCTYPE*	20060624	?
23	\\images\pdf2outputs.jpg	*No DOCTYPE*	20060624	?
24	\\installing\installing_fo.dita	task	20060623	Installing the Idiom FO plugin
25	\\maps\creatingmaps.dita	task	20060623	Creating maps
26	\\images\folderidiomfo.jpg	*No DOCTYPE*	20060622	?
27	\\installing\aboutplugins.dita	concept	20060622	About DITA Open Toolkit plugins
28	\\installing\installing_map.ditamap	map	20060622	?
29	\\installing\installing_plugins.dita	task	20060622	Installing plugins
30	\\installing\linux_installing.dita	concept	20060622	Installing on Linux
31	\\installing\linux_installingant.dita	task	20060622	Installing Ant on Linux
32	\\installing\linux_installingditaot.dita	task	20060622	Installing DITA Open Toolkit on Linux
33	\\installing\linux_installingfop.dita	task	20060622	(Optional) Installing FOP on Linux

ditalinks - URL errors

No debugging help from the Toolkit (but broken links)

Output from ditalinks:

URL `https://sourceforge.net/forum/?group_id=132728` open failed.

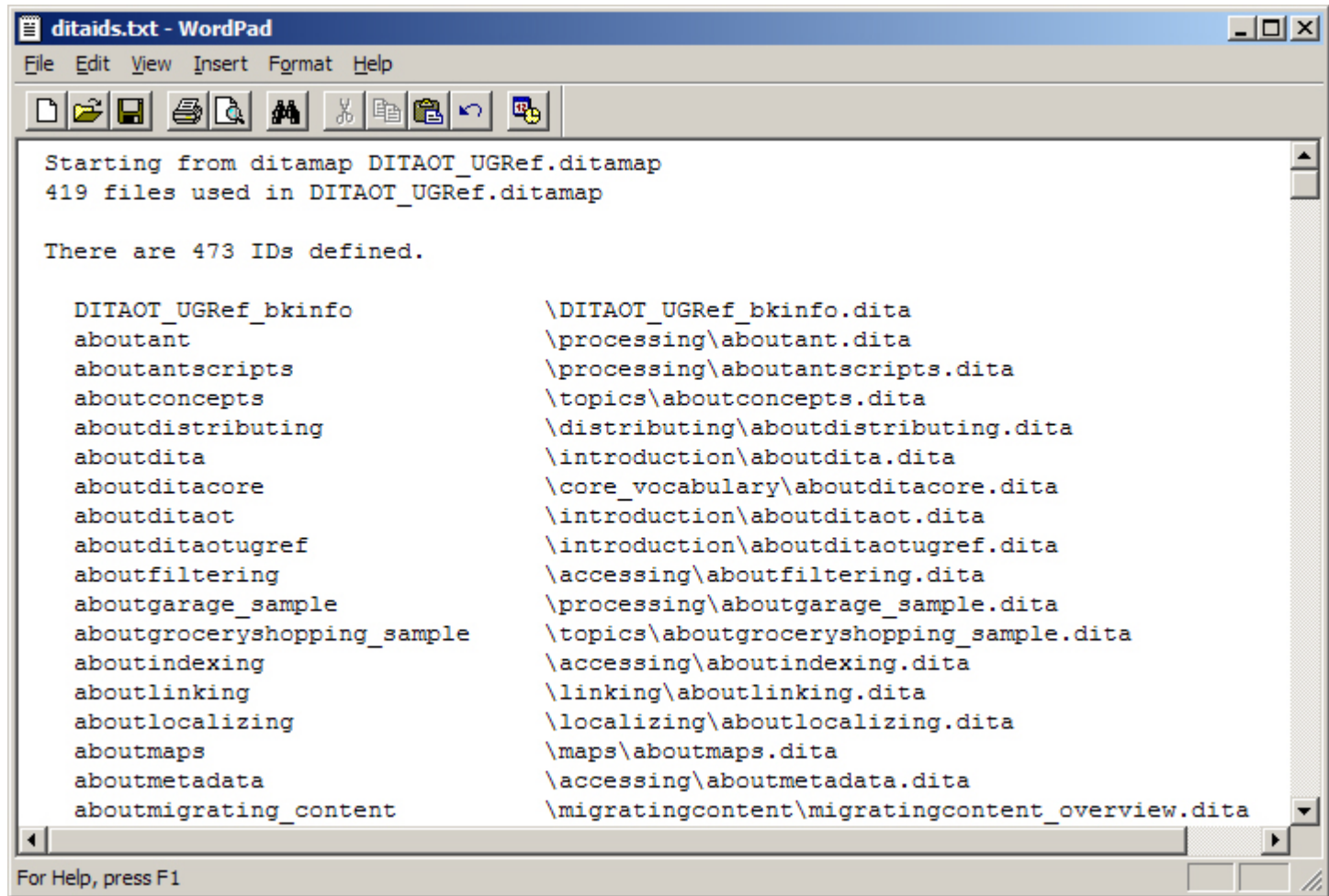
Host `unt.apache.org` invalid for `http://unt.apache.org/bindownload.cgi`.

ditaauthors – author/copyright

```
count  copyright holder
=====
225 VR Communications, Inc.
  4 IBM Corporation
 20 Comtech Services, Inc.
```

```
count  author
=====
179 Anna van Raaphorst
   3 IBM
 45 Richard Johnson
 20 Jen Linton
 20 JoAnn Hackos
 20 Kylene Bruski
   2 Deborah Pickett
```

dit aids – file ID usage



```
Starting from ditamap DITAOT_UGRef.ditamap
419 files used in DITAOT_UGRef.ditamap

There are 473 IDs defined.

DITAOT_UGRef_bkinfo      \DITAOT_UGRef_bkinfo.dita
aboutant                 \processing\aboutant.dita
aboutantscripts          \processing\aboutantscripts.dita
aboutconcepts            \topics\aboutconcepts.dita
aboutdistributing        \distributing\aboutdistributing.dita
aboutdita                \introduction\aboutdita.dita
aboutditacore            \core_vocabulary\aboutditacore.dita
aboutditaot              \introduction\aboutditaot.dita
aboutditaotugref         \introduction\aboutditaotugref.dita
aboutfiltering           \accessing\aboutfiltering.dita
aboutgarage_sample       \processing\aboutgarage_sample.dita
aboutgroceryshopping_sample \topics\aboutgroceryshopping_sample.dita
aboutindexing            \accessing\aboutindexing.dita
aboutlinking             \linking\aboutlinking.dita
aboutlocalizing          \localizing\aboutlocalizing.dita
aboutmaps                \maps\aboutmaps.dita
aboutmetadata            \accessing\aboutmetadata.dita
aboutmigrating_content   \migratingcontent\migratingcontent_overview.dita
```

For Help, press F1

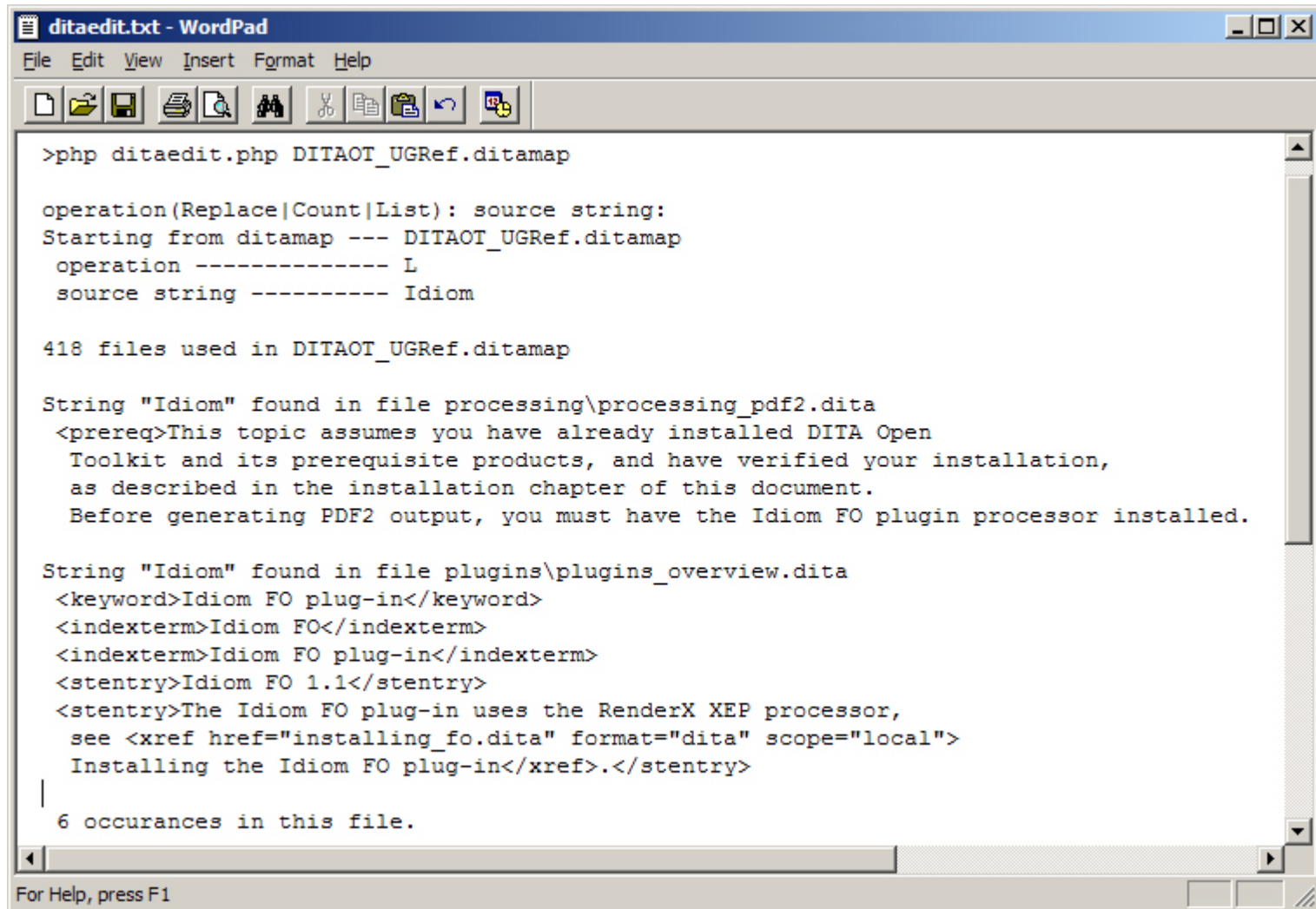
ditakeys - keyword metadata

```
Starting from ditamap DITAOT_UGRef.ditamap
419 files used in DITAOT_UGRef.ditamap

count keyword string                file
=====
  7 accessing information            \accessing\accessing.dita
                                       \accessing\aboutindexing.dita
                                       \accessing\aboutmetadata.dita
                                       \accessing\aboutrdfowl.dita
                                       \accessing\aboutfiltering.dita
                                       \accessing\accessing_production_notes.dita
                                       \accessing\accessing_formoreinfo.dita
  2 administrator audience category \introduction\aboutditaotugref.dita
  2 Adobe FrameMaker                \core_vocabulary\administrator_manager.dita
  2 Adobe FrameMaker                 \installing\installing_required_tools.dita
                                       \migratingcontent\migratingcontent_formoreinfo.dita
  2 Adobe FrameMaker DITA Application P \migratingcontent\migratingcontent_overview.dita
                                       \core_vocabulary\migrating_content.dita
  2 Altova XMLSpy                    \installing\installing_required_tools.dita
                                       \installing\installing_production_notes.dita
```

For Help, press F1

ditaedit – string find/replace



```
>php ditaedit.php DITAOT_UGRef.ditamap

operation(Replace|Count|List): source string:
Starting from ditamap --- DITAOT_UGRef.ditamap
operation ----- L
source string ----- Idiom

418 files used in DITAOT_UGRef.ditamap

String "Idiom" found in file processing\processing_pdf2.dita
<prereq>This topic assumes you have already installed DITA Open
Toolkit and its prerequisite products, and have verified your installation,
as described in the installation chapter of this document.
Before generating PDF2 output, you must have the Idiom FO plugin processor installed.

String "Idiom" found in file plugins\plugins_overview.dita
<keyword>Idiom FO plug-in</keyword>
<indexterm>Idiom FO</indexterm>
<indexterm>Idiom FO plug-in</indexterm>
<stentry>Idiom FO 1.1</stentry>
<stentry>The Idiom FO plug-in uses the RenderX XEP processor,
see <xref href="installing_fo.dita" format="dita" scope="local">
Installing the Idiom FO plug-in</xref>.</stentry>

6 occurrences in this file.
```

For Help, press F1

Production Notes

- We have other experiences in implementing DITA that we can share at a more detailed level
- Contact us for more information (www.vrcommunications.com)
- Sample “production notes” topics:
 - Architecting and planning a DITA solution
 - Creating and using a controlled vocabulary
 - Installing DITA Open Toolkit
 - Setting up your working environment
 - Processing (building) files
 - Troubleshooting the build process
 - Creating topics and maps
 - Linking content
 - Reusing content
 - Including and using metadata
 - Indexing DITA files
 - Using conditional processing (filtering)
 - Customizing your published output
 - Localizing (translating) your DITA content
 - Distributing your published content
 - Migrating legacy content to DITA

(Last Slide) implementing DITA™

For more information:

<http://sourceforge.net/projects/dita-ot/>

> Home Page (for *User Guide*)

www.vrcommunications.com

> *User Guide* source files, tools, sample files

