



Dynamic generation of index entries in DITA source files

Vishal George Palliyathu & Balaji S Kumar
IBM India

Vishal George Palliyathu

- **Senior Information Developer**
- **Member, OASIS DITA TC**
- **Extensive experience in handling multiple middleware transaction processing products**

Balaji S Kumar

- **Senior Engineer**
- **Pioneered innovation at IBM Lotus Sametime**

Part I: A brief overview on Indexing and technical communication

- **Brief history of Indexing**
- **Indexing as a profession**
- **Types of indexing**
- **Indexing at WWW and Google**
- **Why do we need indexes?**
- **Indexing in DITA: Why a good Editor or an IA is also a good Indexer?**
- **Timeline and index optimization**

Brief history of Indexing

- **Printed book indexes appeared in the 1460s, almost from the beginning of the era of printing.**
- **The first printed Biblical concordance was published in 1544; its compiler was burned for heresy.**
- **(What is a concordance? And how is it different from an Index)**
- **- Samuel Johnson's "A dictionary of the English language (1755)" was a first index to the English language.**

Indexing as a profession

- **In the 19th century there were moves to codify indexing. The Index Society was formed in London in 1877 with the aim of creating 'a general index of universal literature'.**
- **Dr Henry Benjamin Wheatley wrote "What is an indexer?" in 1878.**
- **Eventually the Society of Indexers was formed in Great Britain in 1957.**

Types of indexing

- **Bibliographic and database indexing**
- **Genealogical indexing**
- **Geographical indexing**
- **Book indexing**
- **Legal indexing**
- **Periodical and newspaper indexing**
- **Pictorial indexing**
- **Subject gateways**
- **Website and metadata indexing**

Indexing at the WWW and Google

- **The World Wide Web, the largest accumulation of databases, has developed without any overall plan**
- **Few or some sites use sophisticated indexing techniques**
- **Most are only accessible using primitive keyword searches which frequently result in unmanageable numbers of 'hits'. The hits are also not subjected to any kind of quality control with well-documented and spurious information existing side-by-side, along with advertising and other forms of propaganda.**

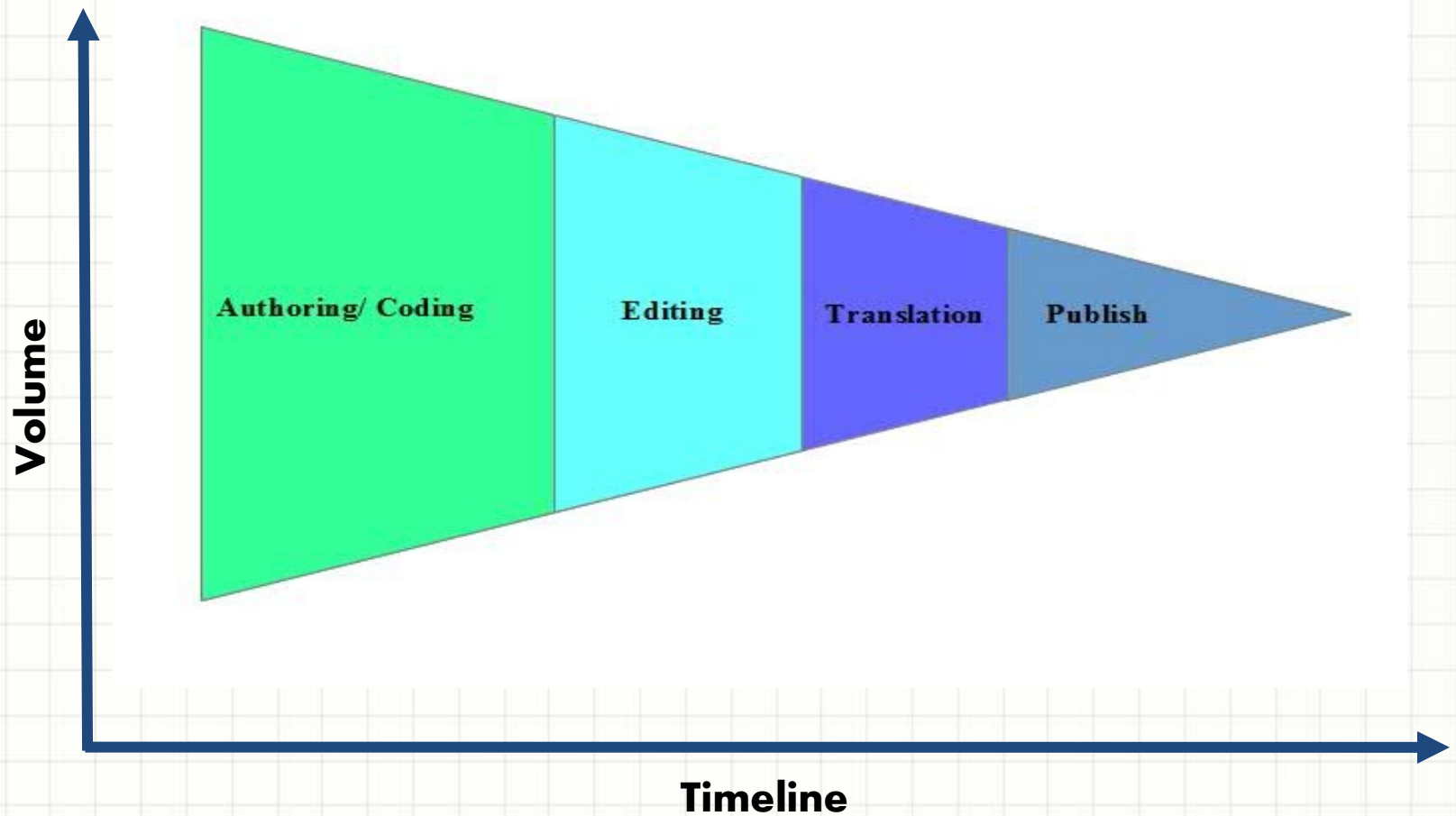
Why do we need indexes?

- **Information needs to be easily findable**
- **Accuracy is a concern**
- **It should be clear and coherent to the point**
- **Some critical business does not have outbound access. Documentation as PDF, with good indexing is imperative.**
- **Progressive disclosure**

Indexing in DITA: Why a good Editor or an IA is also a good Indexer?

- **Indexers organize information, group, classifying data, create taxonomies and develop hierarchies that can be used for navigation.**
- **Use language and words quantitatively and qualitatively to clearly and succinctly describe concepts.**
- **Ensures that users can find information quickly and easily.**
- **Indexers are able to develop information architectures for libraries, and ensure that their information architectures work in conjunction with search engines and thesaurus-controlled vocabularies to provide better retrievability of information.**
- **Indexing is done keeping minimalism and the Nürnberg funnel in mind.**
- **Both professions go hand-in hand, and compliments each other.**

Timeline and index optimization



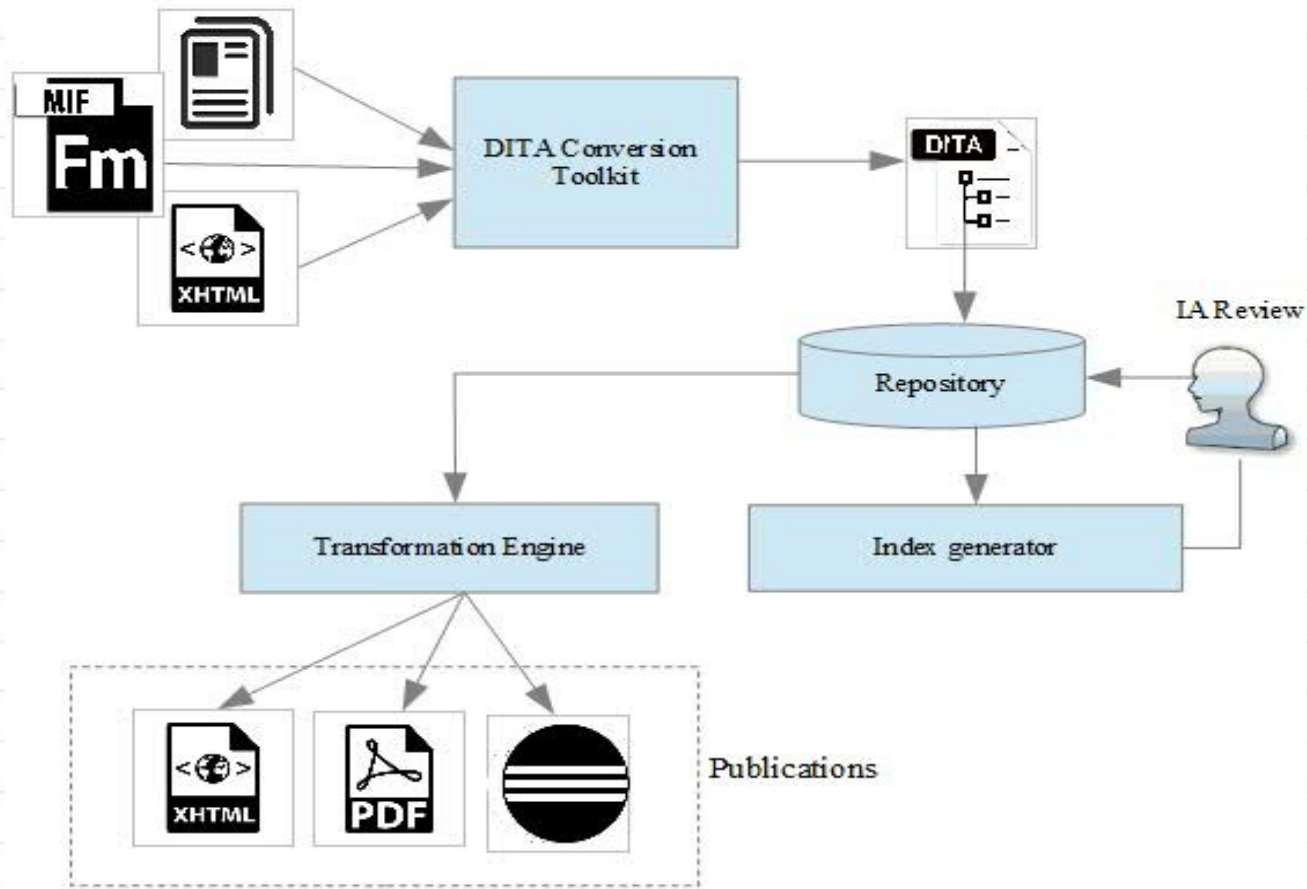
Part II: The need for indexing solutions in modern enterprises

- **Problems and challenges**
- **Solution and requirement**
- **How it works**
- **Plotting the terms and nouns to gerunds and adjectives**
- **Generating the index entries in DITA files**
- **Index harvester**

Problems and challenges

- **Three different products (through acquisition)**
- **Source code in different files formats**
- **Deliverables across multiple mediums**
- **Less time to market**
- **Need to merge terminologies**
- **IA needs to converge**
- **DITA comes to rescue; utility provides a crude index to start with**

Problems and challenges - II



Solution and requirements

Listing the pre-requisite files

- common-symbols.dita
- glossary.dita
- A listing of the enterprise- approved user-goals
- A listing of the commonly used tasks, their gerunds and adjectives

Glossary files

They are DITA files and are created for product libraries.

Contains a brief definition of a technical term

Provides expanded abbreviations

COMMAREA is a technical term defined in the source library

```
<dentry id="x2017636">
  <dt> common work area (CWA) </dt>
  <dd> An area within the CSA that can be used by application programs for user data that needs to be accessed by any task in the system. See also <xref href="#gloss_T/x2203417">
    transaction work area </xref> </dd>
</dentry>
<dentry id="x2198261">
  <dt> communication area (COMMAREA) </dt>
  <dd> A CICS area that is used to pass data between tasks that communicate with a given terminal. The area can also be used to pass data between programs within a task. </dd>
</dentry>
<dentry id="x2051544">
  <dt> communication controller </dt>
  <dd>
  <ol outputclass="glossnoindent">
    <li> 1. A device that directs the transmission of data over the data links of a network; its operation may be controlled by a program executed in a processor to which the controller is connected or it may be controlled by a program executed within the device. (T) </li>
    <li> 2. A type of communication control unit whose operations are controlled by one or more programs stored and executed in the unit. It manages the details of line control and the routing of data through a network. See also <xref href="#gloss_T/x2069228">
      transmission control unit </xref> </li>
  </ol>
</dd>
</dentry>
```

Common symbols

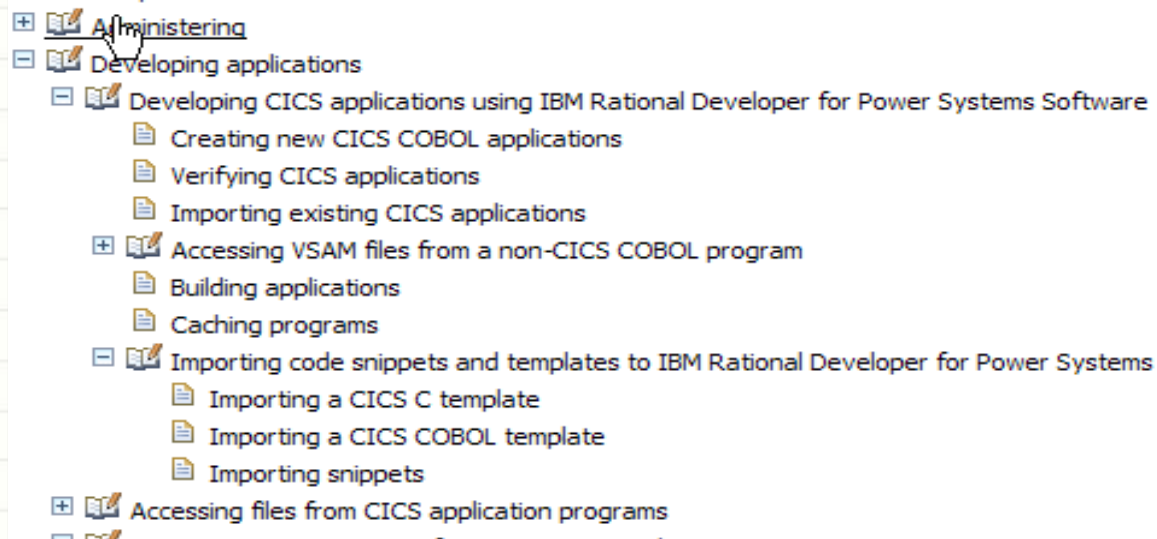
- Are DITA files containing variables and are used throughout the product library
- Usually contains product names, commonly used platforms etc.
- They are mostly nouns
- CICS is the name of a product, widely used across the product library

The screenshot displays a DITA document structure. On the left, a vertical list of product names is shown: AD, Sol, HP, CICS, and IBM. Each name is associated with a 'ord' attribute. The main content area shows a 'body' element containing a 'required-cleanup' attribute and a list of product names with their corresponding keywords and platform identifiers. The 'CICS' entry is highlighted with a black box.

```
body
  required-cleanup remap="text entities" translate="yes" Required cleanup:
  p keyword id="txseries_infocenter" TXSeries Infocenter keyword p
  p keyword id="noncicsproduct" TXSeries for Multiplatforms keyword p
  p keyword id="email" idrcf@hurslev.ibm.com keyword p
  p keyword id="aixos" AIX keyword p
  p keyword id="sunos" Solaris keyword p
  p keyword id="hpos" HP-UX keyword p
  p keyword id="cics" CICS keyword p
  p keyword id="clients" IBM CICS Clients keyword p
```


Gerunds

- *a verbal noun in Latin that expresses generalized or uncompleted action*
- **Tasks performed by users are Gerunds**
- **IBM adheres to User-driven test cases and task-driven documentation**
- **Common tasks performed by a user on a product are all gerunds**



Plotting the nouns and gerunds

- The nouns (from the common-symbols) are listed in column A.
- Terms from the glossary are listed on column B.
- Gerunds are listed in column C.
- Adjectives are listed in column D.

A	B	C	D
Common-symbols	Glossary	Gerunds	A.Gerunds
TXSeries		Installing	Install
TXSeries		Administering	Administer
CICS	region	Creating	Create
CICS	region	Configuring	Configure, configuration
CICS	region	Starting	Start
CICS	region	Stopping	Stop
		Designing	Design
		Tuning	Tune
		Troubleshooting	Troubleshoot
		Communicating	Communication
CICS	Commarea	Migrating	Migrate, migration
		Transferring	Transfers
		Managing	
		Emulating	Emulation
		Modifying	

How it works

- **All the DITA files in the library are checked for the nouns listed in common-symbols. If entries are found, they are listed**
- **The listed entries are then checked against entries in glossary. If entries match, they are listed adjacent to each other.**
- **Gerunds and adjectives are also listed.**
- **Existing index entries, if any are also listed.**
- **A listing of the index entries is generated using the tool. Permutations and combinations of the entries is generated and submitted for review.**

Plotting the terms and nouns to gerunds and adjectives

Common-symbols	Glossary	Gerunds	A.Gerunds	Matched against existing index	Topic that needs to be indexed	Existing index entry	Indexing level			
							Level 1	Level 2	Level 3	Level 4
TXSeries		Installing	Install		Installing TXSeries for Multiplatforms	none				
TXSeries		Administering			Administering TXSeries as a non-root user	none				
CICS	region	Creating			Creating a CICS region	none				
CICS	region	Configuring	Configure, configuration		Configuring and modifying CICS regions	none	Configuring	CICS	Region	
CICS	region	Starting		NA	Starting a CICS region	none	Starting	CICS	Region	
CICS	region	Starting		NA	Starting a CICS region using the cicscp command	none	Starting	CICS	Region	(cicscp command)
CICS	region	Starting		NA	Starting a CICS region using the Administration Console	none	Starting	CICS	Region	(administration console)
CICS	region	Stopping		NA	Stopping a CICS region	none	Stopping	CICS	Region	
CICS	region	Stopping		NA	Stopping a CICS region using the cicscp command	none	Stopping	CICS	Region	(cicscp command)
CICS	region	Stopping		NA	Starting a CICS region using the Administration Console	none	Stopping	CICS	Region	(administration console)
		Designing								
		Tuning								
		Troubleshooting								
		Communicating	Communication							
CICS	Commarea	Migrating	Migrate, migration	XCTL	Migrating XCTL commands that pass COMMAREAs		Migrating	XCTL	Commarea	
		Transferring	Transfers		XCTL					
		Managing								
		Emulating	Emulation							

Generating the index entries in the DITA files

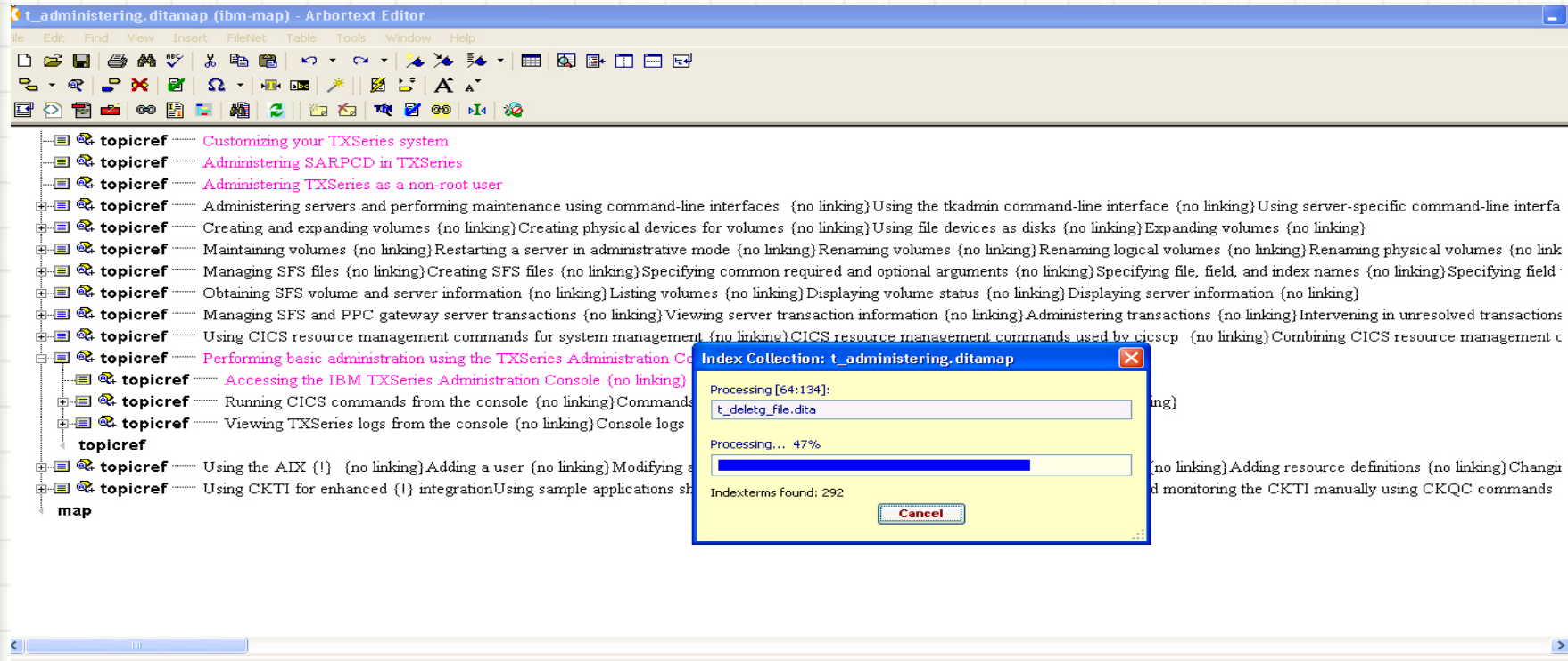
The screenshot shows a DITA editor window with a document titled "Migrating XCTL commands that pass COMMAREAs". The document structure includes a title, prolog, keywords, and a taskbody. The keywords section contains several index terms: "migrating", "XCTL", and "COMMAREAs". The taskbody contains a section titled "About this task" and a table with two columns: "Before" and "After".

The table compares the code for a program named "PROG1" before and after migration. The "Before" column shows the original code, and the "After" column shows the migrated code.

Program	Before	After
PROG1	<pre>EXEC CICS XCTL PROGRAM(PROG2) COMMAREA (structure)</pre>	<pre>EXEC CICS PUT CONTAINER (structure-name) CHANNEL (channel-name) FROM (structure) EXEC CICS XCTL PROGRAM(PROG2) CHANNEL (channel-name)</pre>

Index harvester

- The results can be harvested through the index harvester



Command:

Index harvester - II

- The index harvester captures the index entries and provides listing in an editable UI, for further optimization.

Edit Index Collection: t_administering-Index.ditamap

File View Find Options Tools Window Help

Index Entry [20 of 354]

i1: CICS regions i2: starting i3:

File Reference: t_strt_rgn_commandline.dita

Update Row Insert Row Delete Row Undo Row Cancel Changes

Status	i1	i2	i3	i4	Topic Title	File Name	Map Name	T...
	procedures	SFS	starting		Starting an SFS server by using the cicssfs command	t_strt_sfs_srvr_cicssfs.dita	t_administering.ditamap	task
	procedures	configuring SFS			Setting up the queues by using the cicssfsconf command	t_setg_queues_cicssfsconf.dita	t_administering.ditamap	task
	procedures	configuring a region			Using the cicdefault command to define a region	t_usg_cmd_define_rgn.dita	t_administering.ditamap	task
	procedures	creating SFS logical volumes			Preparing the physical storage devices for the SFS server	t_prepq_storage_devices_sfs_srvr.dita	t_administering.ditamap	task
	querying	OFDs			Querying an open file descriptor	t_querying_open_file_descriptor.dita	t_administering.ditamap	task
	querying	SFS information			Displaying server information	t_displayg_servr_info.dita	t_administering.ditamap	task
	querying	file locks in SFS			Displaying file locks	t_displaying_file_locks.dita	t_administering.ditamap	task
	querying	files in SFS			Querying a file	t_querying_file.dita	t_administering.ditamap	task
	querying	transaction information			Displaying transaction information	t_displaying_trans_info.dita	t_administering.ditamap	task
	RD attributes, use of	StartType			Starting a region from the command line	t_strt_rgn_commandline.dita	t_administering.ditamap	task
	RD attributes, use of	StartType			Starting a CICS system	t_strt_tx_system.dita	t_administering.ditamap	task
	reclaiming	disk space			Reclaiming storage space	t_reclaimg_storage_space_nonaixwinnt.dita	t_administering.ditamap	task
	reclaiming	disk space (AIX)			Reclaiming storage space (AIX logical volumes)	t_reclaimg_storage_space_aix.dita	t_administering.ditamap	task
	records	specifying field sizes in SFS files			Specifications for field types and sizes	t_specifying_field_types_and_sizes.dita	t_administering.ditamap	task
	records	specifying field types in SFS files			Specifications for field types and sizes	t_specifying_field_types_and_sizes.dita	t_administering.ditamap	task
	records	updating (figure)			Creating entry-sequenced files	t_creatg_entry-sequenced_files.dita	t_administering.ditamap	task
	recoverable temporary storage queues				Setting up the queues by using sfsadmin commands	t_setg_queues_sfsadmin.dita	t_administering.ditamap	task
	region	cicdefault command			Using the cicdefault command to define a region	t_usg_cmd_define_rgn.dita	t_administering.ditamap	task
	region	configuring			Using the cicdefault command to define a region	t_usg_cmd_define_rgn.dita	t_administering.ditamap	task
	region	default user ID			Creating the region's default user ID	t_crt_rgn_default_userid.dita	t_administering.ditamap	task
	region	shutdown			Shutting down the region by using cicscp stop region co...	t_sht_rgn_usg_cicscp_stop_rgn.dita	t_administering.ditamap	task
	region	shutdown			Shutting down the region by using EXEC CICS PERFOR...	t_sht_rgn_usg_execicperform.dita	t_administering.ditamap	task
	region	shutdown			Shutting down the region by using CEMT PERFORM SHU...	t_sht_rgn_usg_cemtperformshutdown.dita	t_administering.ditamap	task
	region	starting			Starting a region from the command line	t_strt_rgn_commandline.dita	t_administering.ditamap	task
	region	starting			Starting a CICS region using the TXSeries Administratio...	t_strt_tx_rgn.dita	t_administering.ditamap	task
	region	startup			Starting the region by using the cicscp start region com...	t_strt_rgn_cicscp_start_sfs_srvr_cmd.dita	t_administering.ditamap	task
	Region Definitions (RD)	overriding at startup			Starting a CICS system	t_strt_tx_system.dita	t_administering.ditamap	task
	Region Definitions (RD)	overriding at startup			Starting a region from the command line	t_strt_rgn_commandline.dita	t_administering.ditamap	task
	Region Definitions (RD) attributes, use of	StartType			Starting a CICS system	t_strt_tx_system.dita	t_administering.ditamap	task
	Region Definitions (RD) attributes, use of	StartType			Starting a region from the command line	t_strt_rgn_commandline.dita	t_administering.ditamap	task
	relative files	creating			Creating relative files	t_creatg_relative_files.dita	t_administering.ditamap	task
	renaming	AIX logical volumes			Renaming AIX logical volumes	t_renamg_aix_vols.dita	t_administering.ditamap	task
	renaming	files in SFS			Renaming a file	t_renamg_file.dita	t_administering.ditamap	task



Questions?