PlayShakespeare.com XML Specification

Category	Element	Description	Required Attributes	Optional Attributes	Description	Example
Overall Structure	<play>	The root level wrapper for all plays and must contain both the <title> and <titleshort> elements.				
	<title>	The full title of element, which may be a play or section.				
	<titleshort>	The short title of element, which may be a play or section.				
	<subtitle>	Used for secondary titles in the poems.				
	<edition>	The edition of this work.				
	<playwright>	Used within the <play> or <poem> tag to indicate the author of the work.				
	<pubdate>	The date (or date range) the work was originally published.				
	<act>	Each play act is wrapped in this tag.	num		The unique number for this act within the play.	<act num="3">
	<acttitle>	The title of the play act. Usually "Act 3".				
	<actfinis>	Used within the <act> to identify a finis closing statement to close the act.				
	<prologue>	Used within the <act> on the same level as <scene>, it identifies a speech by an individual character to open an act or the play.				
	<prologuetitle>	Currently not used.				
	<epilogue>	Used within the <act> on the same level as <scene>, it identifies a speech by an individual character to close an act or the play.				
	<epiloguetitle>	Currently not used.				
	<finis>	The container tag to wrap any finis information at the end of a work.				
	<finistitle>	Used within <finis> and identifies the finis statement at the end of a work.				
Scenes	<scene>	Each scene within an act is wrapped in this tag. Scenes are always wrapped within the <act> tag.		actnum	The unique number for the containing act within the play (same as the num attribute of <act>).	<scene actnum="3" num="1" numberOfLines="58">
				num	The unique sequential number for this scene within the act.	
				numberOfLines	A count of the numebr of lines within this scene.	
	<scenetitle>	The title of the play scene. Usually "Scene 1".				
	<scenelocation>	The location of the scene.				
	<scenepersonae>	The full names of the characters within a scene. This tag must contain at least one character in a <scenepersona> tag.				
	<scenepersona>	The full name of a character within a scene.		short	The shortened name of a character.	<scenepersona short="1. GENT.">First Gentleman</scenepersona>
Scene Descriptions	<scenedescription>	Identifies the description of the action within a scene.				
	<poemdescription>	Identifies the description of the action within a poem.				
	<lineref>	Used to indicate a line reference in a poem description.				
	<descriptions>	Currently not used.				
Dialogue	<speech>	The wrapper for character speeches. It must include only one <speaker> tag and at least one <line> tag.				
	<song>	The wrapper for character songs. It must include only one <speaker> tag and at least one <line> tag.				
	<speaker>	The short name of the speaker.		long	The long name of the character	<speaker long="Earl of Gloucester">GLOU.</speaker>
				alias	The former name of a character that has changed during the course of the play or to indicate multiple characters are speaking at once.	<speaker alias="L. GREY.">Q. ELIZ.</speaker>
	<line>	The line(s) spoke by the speaker.		globalnumber	The unique global number for this line in this play.	<line globalnumber="1322" number="268" form="verse" offset="4" order="first">Followed the old man forth.</line>
				number	The unique number for this line in this scene.	
				form	Values include verse, prose, rhyme, or lyric and indicate the poetic form of the line.	

				offset	Values include 0-9 and used to identify short lines are shared lines by the number of syllables spoken before that line begins.	
				lyric	Used to identify when a character is singing or incanting.	
				order	The first lines of a speech to indicate formatting with a dropcap (e.g. "first", "second", "third", "fourth").	
	<recite>	Identifies a line or passage when a character is reading or reciting.				
Stage Directions	<stagedir>	Stage directions to indicate character action. Directions may be inside a character's <speech> (not inside the <line> tag) or in between different character lines (within a <scene>). This must contain the <dir> tag to show the direction.				<stagedir> <dir type="action">Gives her the letter.</dir> <actor long="Orlando">ORL.</actor> <recipient long="Rosalind">ROS.</recipient> </stagedir>
				sdglobalnumber	The unique number for this direction comprised of the previous global line number and a two-digit sequential number after the decimal place.	<stagedir sdglobalnumber="1322.01" sdnumber="268.01">
				sdnumber	The unique number for this direction comprised of the previous scene line number and a two-digit sequential number after the decimal place.	
	<dir>	The actual stage direction text from the play.		type	Values include: enter, exit, exeunt, aside, action, read, sound, fight, speak, song, death. The "exeunt" tags indicate all characters exit unless a character name is within the <actor> tags. Any <actor> tag with "ALL." in it indicates all characters currently present in a scene perform that action.	<dir type="enter">Enter Gloucester.</dir> <dir type="exeunt">Exeunt.</dir> <dir type="exit">Exit Douglas.</dir> <dir type="death">Douglas kills Blunt.</dir> <dir type="fight">They fight.</dir> <dir type="action">He throws the bottle at him.</dir>
	<actor>	Used to indicate the active doer of the action in a stage direction. May only be contained within the <stagedir> tag.		short		<actor short="RUM.">RUM.</actor>
				long		<actor long="Rumor">RUM.</actor>
	<recipient>	Used to indicate the passive recipient of the action in a stage direction. May only be contained within the <stagedir> tag.		short		<recipient short="RUM.">RUM.</recipient>
				long		<recipient long="Rumor">RUM.</recipient>
Dramatis Personae	<personae>	Identifies the dramatis personae at the beginning of a play.				
	<perstitle>	Currently not used.				
	<persnamestitle>	Currently not used.				
	<persona>	Identifies and individual character within the <personae> tag.	gender		Used to indicate whether the character is male or female.	<persona gender="male">
	<persname>	Identifies the character's name within the <personae> tag. May also be used in the <persaliases> tag to indicate alternate names for a given character.		short		<persname short="KING." numberOfLines="373" numberOfVerseLines="366" numberOfProseLines="7" numberOfLyricsLines="0">King of France</persname>
				numberOfLines		
				numberOfVerseLines		
				numberOfProseLines		
				numberOfLyricsLines		
	<persscenes>	Used to identify the collection of scenes a character is in.		numberOfScenes		<persscenes numberOfScenes="5">
	<persscene>	Used to indicate an act/scene number a character is in. It must appear within a <perscenes> tag. Uses dot notation "act.scene" or scene title.				
	<persaliases>	Used to identify one or more alternate names for the character. It must reside within a <persona> tag.				
	<persdescription>	Used to identify the description of a character.				
	<persnameref>	A reference to a character within a character description.				
	<perstitleref>	A reference to another work within a character description.				
Poems	<poem>	The root level wrapper for all poems and must contain both the <title> and <titleshort> elements.				
	<poemquarto>	Currently not used.				

Category	Tag	Description			Attribute Description	Example
	`<poembody>`	Identifies the main part of a poem after the introduction/front matter.				
	`<poemintro>`	The introduction or front matter of a poem.				
	`<intro>`	A subsection of <poemintro> used to identify an introductory phrase.				
	`<dedication>`	A subsection of <poemintro> used to identify a dedication.				
	`<dedbody>`	Currently not used.				
	`<dedtitle>`	Currently not used.				
	`<dedsubtitle>`	Currently not used.				
	`<dedauthor>`	A subsection of <dedication> used to identify the author/signature of a dedication.				
	`<dedclosing>`	Currently not used.				
Sonnets	`<sonnet>`	Used to identify a single sonnet.				
	`<sonnetnum>`	A unique and sequential numeric value assigned to a sonnet. It muct reside within the <sonnet> tag.				
	`<argument>`	Currently not used.				
	`<stanza>`	Identifies a collection of rhyming lines as a division of a poem.		num		`<stanza num="7">`
	`<stanzasmall>`	Identifies a collection of rhyming lines as a division of a poem. (Should be merged with <stanza>.)				
	`<stanzanum>`	The number of the current stanza displayed to the reader.		num		`<stanzanum num="7">VII</stanzanum>`
	`<couplet>`	Used to indicate a pair of <line> elements.				
	`<quatrain>`	Used to identify four lines of rhyming poetry in <line> elements.		quanum	The unique and sequential ID of the quatrain.	`<quatrain quanum="3">`
	`<missing>`	Used to identify content that is presumed missing from the works.				
Folio & Quartos	`<endpersonae>`	Used to identify the dramatis personae at the end of the folio/quarto editions.				
	`<personagroup>`	A group of characters in the dramatis personae.	title		The title describing a subset of characters (each wrapped in a <persona> tag) in the dramatis personae list.	`<personagroup title="Of the Kings Partie." items="6">`
				items	The number of <persona> items in a <personagroup>	
	`<persnames>`	A subtitle within the <endpersonae> to state the names of the actors.				
	`<persloc>`	Location information in the dramatis personae.				
	`<personagrouptitle>`	Not currently in use.				
	`<actortitle>`	Not currently in use. Likely duplicate of <persnames>.				
Unique/Foreign Words and Formatting	`<latin xml:lang="la">`	Language tags are used to indicate source of foreign language words used.				
	`<french xml:lang="fr">`					
	`<italian xml:lang="it">`					
	`<dutch xml:lang="nl">`					
	`<spanish xml:lang="es">`					
	`<dropcap>`	Identify a dropcap character at the beginning of a speech.		type	Indicates the type of dropcap used (e.g. "floral").	`<dropcap type="floral">A</dropcap>`
	`<titleref>`	The formal name of the title of a work in the spoken text (e.g the title of a play).				
	`<nameref>`	The formal name of an person in the spoken text.				
	`<italic>`	Apply an italic font style the the contents.				
	`<noitalic>`	Remove and italic font style from the contents.				
	`<lb />`	Manual line breaks where indicated in the source material.				
	`<specialq>`	Currently not used. Used to identify specially formatted Q characters.				
Text/Play Sources	`<sources>`	Used to identify a list of works Shakespeare used as sources for the work.				
	`<playsource>`	Used within the <sources> container to wrap an individual work/source.				
	`<playsourcetitle>`	The title of a work used as source material. Must be used within a <playsource> tag.				

	<playsourceauthor>	The author of a work used as source material. Must be used within a <playsource> tag.			
	<playsourcedate>	The publication of a date of a work used as source material. Must be used within a <playsource> tag.			
	<textsources>	Used within the <sources> container to wrap an individual source material from Shakespeare in order to create the edition.			
	<textsource>	Used within the <textsources> container to wrap an individual Shakespeare work.	title	The title of the source (e.g "Folio", "Quarto", "Manuscript").	<textsource title="Folio" edition="1" date="1623">
			edition	The numerical edition number.	
			date	The publication date of the text source edition.	
	<maintitle>	Currently not used.			
File Source/License	<sourcedetails>	The license and author information is at the bottom of each play. This information may be hidden using XSL/CSS, but may not be removed.			
	<source>	The copyright owner of the material.			
	<sourceurl>	The source URL as designated by the copyright holder.			
	<copyright>	The copyright notice.			
	<version>	The version of the file.			
	<license>	The license under which the work is covered.			
	<licenseurl>	The source URL for the license.			
	<termsurl>	The terms for distribution and usage.			
	<notice>	Currently not in use.			