



中华人民共和国国家标准

GB/T XXXXX—XXXX

信息技术 流式文档互操作性的度量

Information technology - Measurement of flow document interoperability

点击此处添加与国际标准一致性程度的标识

(征求意见稿)

在提交反馈意见时，请将您知道的相关专利连同支持性文件一并附上

XXXX - XX - XX 发布

XXXX - XX - XX 实施

中华人民共和国国家质量监督检验检疫总局
中国国家标准化管理委员会 发布

前 言

本标准按照GB/T 1.1—2009给出的规则起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别这些专利的责任。

本标准由全国信息技术标准化技术委员会（SAC/TC28）提出并归口。

本标准起草单位：北京信息科技大学，中国电子技术标准化研究院，北京金山办公软件有限公司，无锡永中软件有限公司，中标软件有限公司，航天福昕软件（北京）有限公司。

本标准主要起草人：侯霞、李宁、吴新松、方春燕、张展新、冯辉，王耀、荣明军、史周波。

引 言

GB/T AAAA-XXXX《电子文件存储与交换格式 流式文档》中规定了中文办公软件流式文档格式规范，此外目前还存在着多种广为使用的流式文档格式。为了准确衡量使用不同格式的文档之间的互操作性，提高文档在不同的应用系统中的共享和交换能力，特制订本标准。

本标准通过文档互操作性度量模型（简称度量模型）度量文档的互操作性。该模型包含文档的一般抽象概念及其相互关联，并记录这些概念对应的互操作属性。一个具体的文档实例的格式可以与该模型的概念体系进行映射，并依据相应的互操作属性计算得到文档的互操作性。由于本标准的度量模型基于GB/T AAAA-XXXX，文档的互操作度量结果直观反映了某文档实例与GB/T AAAA-XXXX的互操作性，借助该度量结果进而可以获得任何文档之间的互操作性。

信息技术 流式文档互操作性的度量

1 范围

本标准规定了度量流式文档互操作性所使用的模型和方法。

本标准适用于文档的编写者和使用者在文档生存周期中对文档的创建、编辑、修改和评阅等活动。

2 规范性引用文件

下列文件对于本文件的应用是必不可少的。凡是注日期的引用文件,仅注日期的版本适用于本文件。凡是不注日期的引用文件,其最新版本(包括所有的修改单)适用于本文件。

GB/T AAAA-XXXX 电子文件存储与交换格式 流式文档

ISO/IEC 13250 信息技术 文档描述与处理语言 主题图 (Information Technology-Document Description and Processing Languages-Topic Maps)

3 术语、定义和缩略语

3.1 术语和定义

下列术语和定义适用于本文件。

3.1.1

功能点 feature

文档格式的功能单元,构成文档格式的整体功能。

3.1.2

流式文档 flow document

按照逻辑关联描述文档结构,显现时按灌排方式临时计算版面的一类文档。

3.1.3

文档格式 document format

文档内容与显现式样的组织形式。文档格式一般遵循某种标准。

3.1.4

文档互操作性 document interoperability

一种格式的文档转换为另一种格式的能力。

3.1.5

文档实例 document instance

某个具体的文档,用以区别对文档的统称。

3.1.6

主题 topic

Topic Map中的知识概念。

3.1.7

主题图 topic map

一种用于描述信息资源的知识结构的元数据形式，用于表示知识概念间的相互联系，以及定位某一知识概念所在的资源位置。

3.2 缩略语

下列缩略语适用于本文件。

DI 功能点重要度 (Degree of Importance)

DT 功能点可转换度 (Degree of Transformability)

ODF 面向办公应用的开放文档格式 (Open Document Format for Office Applications)

OOXML 开放办公文档 XML 格式 (Office Open XML File Formats)

UOF 统一办公文档格式 (Uniform Office Document Format)

XML 可扩展置标语言 (Extensible Markup Language)

XTM XML 表示的主题图 (XML Topic Map)

4 互操作性度量模型

4.1 概述

本标准的互操作性度量模型用ISO/IEC 13250规定的主题图来表示。

度量模型结构的示意图如图1所示，分为概念层和资源层，具体说明见4.2和4.3。概念层描述功能点的概念及其相互关系，资源层包括特定的文档格式信息及其互操作性信息。

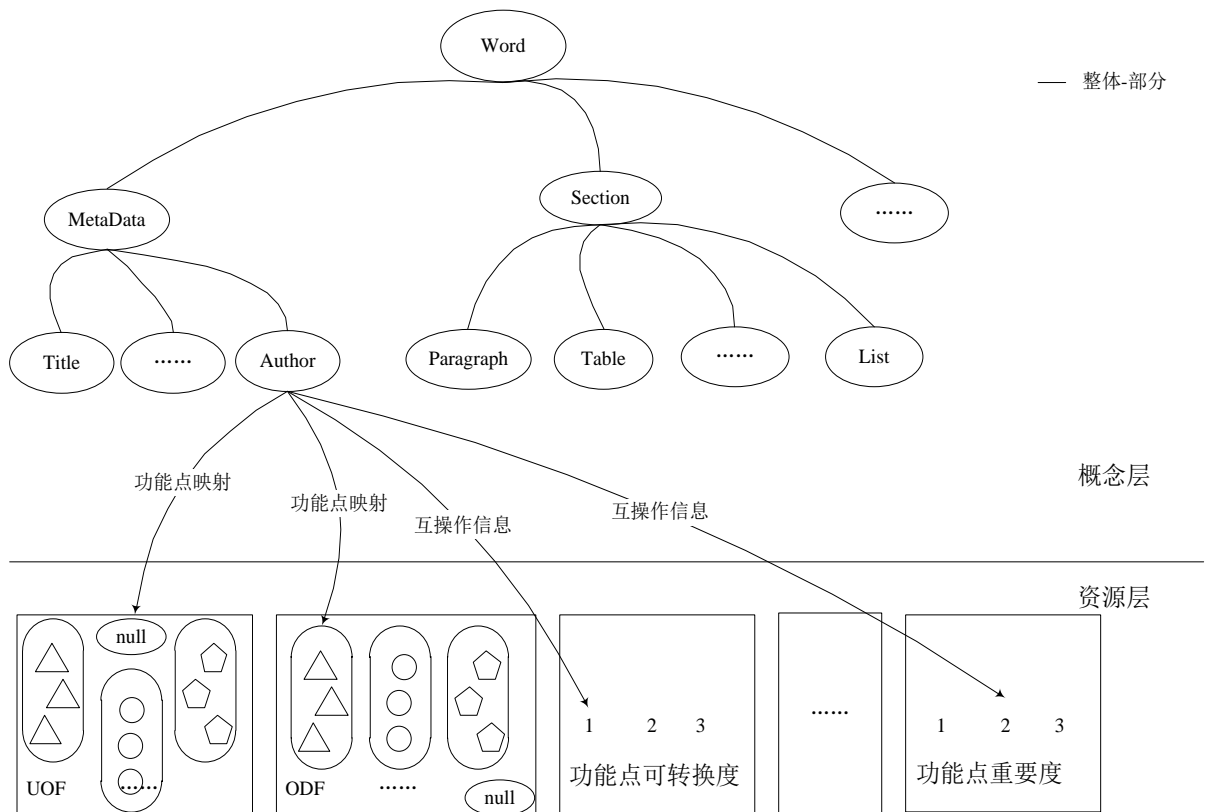


图 1 文档互操作性度量模型

4.2 概念层

概念层中的概念体系表示为节点和边构成的树形结构。节点为格式功能点对应的概念，也称为主题（Topic）。例如，段落（Paragraph）和句（Run）。边表示概念之间的关联，一般为整体-部分（whole-part）关联。例如，图1概念层的树型结构描述的是字处理文档的示意图，实线表示“整体-部分”关系，用于描述文档中各部分之间的层次关系。其中，功能点“字处理文档”（Word）包含子功能点“元数据”（Metadata）、“章节”（Section）等；功能点“元数据”包含子功能点“题目”（Title）和“作者”（Author）等；功能点“章节”包含子功能点“段落”（Paragraph）、“表格”（Table）和“列表”（List）等。

可以根据需要引入其他的关联类型。

4.3 资源层

4.3.1 概述

资源层描述用于度量互操作性的信息。模型中只为概念树中的叶节点指定资源信息。资源包含以下三个属性：所属类型（Type）、有效范围（Scope）和资源值（Value）。

资源按照其描述信息的用途分为两类：

- a) 映射资源：描述概念体系与具体文档格式的映射关系。
- b) 度量资源：描述互操作性度量参数。

4.3.2 映射资源

映射资源指出概念在某个文档格式中的位置信息，由以下三个属性进行描述：

- a) 所属类型：表示特种特定的文档格式，如 Type=UOF、Type=OOXML 或 Type=ODF。
- b) 有效范围：表示功能点在“所属类型”所指文档格式中所处的物理文件位置。
- c) 资源值：表示功能点在“有效范围”所指物理文件中的具体位置。对于文档格式不支持的功能点，可用 Value = null 表示。对于采用 XML 描述的文档格式而言，资源值宜用 XPath 来描述。

大多数文档格式都采用了打包的形式，即一个文档由若干物理文件组成。此时，“有效范围”用来指定功能点所对应元素所在的物理文件在包中的相对位置及文件名。例如，对于图1中功能点“作者”（Author），其Type=“UOF”，Scope=“/_meta /mtea.xml”，Value=“/元:元数据_5200/元:作者_5204”，说明该功能点在UOF格式中，位于物理文件“mtea.xml”，在其中具体位置的XPath为“/元:元数据_5200/元:作者_5204”。对于仅采用单文件而不采用打包形式的文档格式，Scope=“.”。

4.3.3 度量资源

度量资源表示互操作性相关的信息。对应于概念层的每个功能点，有以下两个互操作参数：

- a) 功能点重要度（DI），即功能点在文档格式中的重要程度。该值一般根据文档功能点的使用频度确定。功能点重要度同样使用所属类型、有效范围和资源值三个属性来描述：
 - 1) 所属类型：Type=“DI”；
 - 2) 有效范围：不使用；
 - 3) 资源值：取值为[0,1]，[0-0.3]表示该功能点很少使用，(0.3-0.6)表示常用功能点，[0.6-1.0]表示必备的功能点。
- b) 功能点可转换度（DT）：即某个功能点在不同格式中可转换的程度。该值一般是通过对文档格式规范分析后确定的经验值。可转换度同样使用所属类型、有效范围和资源值三个属性来描述：
 - 1) 所属类型：Type=“DT”；

- 2) 有效范围：一般的形式是 A,B，表示由格式 A 转换到格式 B；
- 3) 资源值：取值为[0,1]， [0-0.3]表示该功能点无法转换或很难转换， (0.3-0.6)表示转换难度中等， [0.6-1.0]表示容易转换。

注：功能点的可转换度不一定是对称的，即某功能点从格式A转换到格式B的可转换度可能不等于从格式B到格式A的可转换度。

5 互操作性度量方法

根据度量模型中的参数，文档从格式A转换到格式B的互操作性度量公式为：

$$\text{Interop}(A, B) = \frac{1}{|T_A|} \sum_{i=1}^{|T_A|} O^{DI}(t_{Ai}) O_{A,B}^{DT}(t_{Ai}) \dots\dots\dots (1)$$

式中：

T_A ——文档系统 A 用到的功能点（主题）集合。 $|T_A|$ 为 T_A 中功能点的数量， $|T_A|>0$ 。

$t_{Ai} \in T_A, 1 \leq i \leq |T_A|$ —— T_A 的第 i 个功能点。

$O^{DI}(t_{Ai}) \in [0,1]$ ——即为 4.3.3 中所述功能点 t_{Ai} 的重要度的资源值（Value），其 Type=“DI”。

$O_{A,B}^{DT}(t_{Ai}) \in [0,1]$ ——从格式 A 转换到格式 B 时，功能点 t_{Ai} 的可转换度的值。 $O_{A,B}^{DT}(t_{Ai})$ 即为 4.3.3 中所述功能点 t_{Ai} 的可转换度的值（Value），其 Type=“DT”，Scope=“A,B”。

$\text{Interop}(A, B)$ 的值在[0,1]区间，值越大表示互操作性越好。

在无法直接度量两个文档实例（如A，C）的互操作性时，可以度量经由中间格式的互操作性。例如：当已知 $\text{Interop}(A, B)$ 和 $\text{Interop}(B, C)$ 时， $\text{Interop}(A, C) = \text{Interop}(A, B) \times \text{Interop}(B, C)$ 。

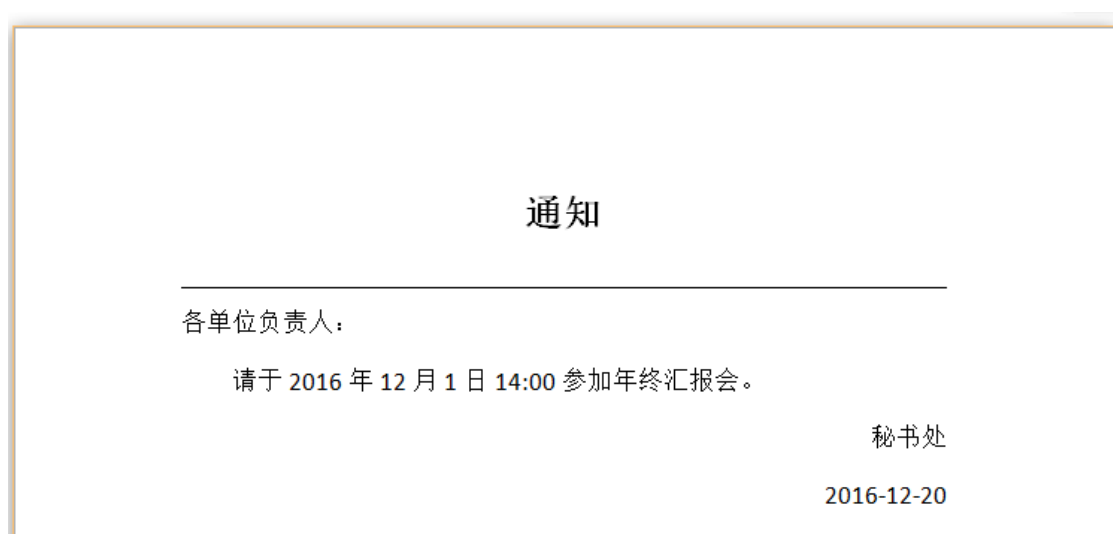
将 T_A 替换成文档格式标准涉及的所有功能点集合时，上述公式可以计算两种文档格式标准之间的互操作性。

6 互操作性度量的应用

文档应用中可以根据需要设定互操作性的阈值。当互操作性小于阈值时，应该调整文档实例中互操作性差的功能点，代之以互操作性较好的功能点，以提高文档整体的互操作性。

附 录 A
(资料性附录)
一个模型应用的例子

本附录给出一个例子，说明如何使用度量模型。假设用户用OOXML格式写一个文档（如图A.1所示），其他用户使用支持ODF格式的文档编辑器打开它。实际上，用户可以使用OOXML中的不同功能点集合来设计一些看起来相同的文档。但这些文档从OOXML到ODF的互操作性却不尽相同。用户应该使用互操作性高的文档，以便于信息交换。



图A.1 一个实例文档的视图

假设已建立好一个满足第4章所述要求的文档互操作性度量模型。基于该度量模型，用户可以根据度量结果从多个备选方案中选择互操作性高的文档。

例如，用户设计了两个文档，分别称为实例1和实例2，它们使用的功能点集合不同，如表A.1所示。表中各项说明为：

“功能点 f_i ”列：度量模型中的功能点名称。

“OOXML”列：指明OOXML格式是否支持功能点 f_i 。“1”表示“支持”，“0”表示不支持。在“ODF列”中类似，指明ODF格式是否支持该功能点。

“实例1”列：指明实例1中是否使用了功能点 f_i 。“1”表示“使用”，“0”表示未使用。“实例2”列类似，指明实例2中是否使用了该功能点。

“ $O_{OOXML,ODF}^{DT}(f_i)$ ”列：给出功能点从格式OOXML转换到格式ODF时，功能点 f_i 的可转换度的资源值。

“ $O^{DI}(f_i)$ ”列：给出功能点 f_i 的重要度的资源值。

度量参数 $O_{OOXML,ODF}^{DT}(f_i)$ 和 $O^{DI}(f_i)$ 在模型的资源层中描述，说明见4.3.3。使用公式(1)度量文档

实例1和实例2的互操作性， $Interop(OOXML,ODF)$ 值的计算结果为：实例1为0.907，实例2为0.926。由此可知，实例2的互操作性较高。

表A.1 文档格式与实例的互操作信息

序号	功能点 f_i	OOXML	ODF	$O_{OOXML,ODF}^{DT}(f_i)$	$O^{DI}(f_i)$	实例 1	实例 2
1	Abstract	1	1	3	3	1	1
2	BottomMargin	1	1	3	3	1	1
3	CharactersCount	1	1	3	2	1	1
4	CreateDate	1	1	3	3	1	1
5	CreatedByApp	1	1	3	3	1	1
6	DocTemplate	1	1	3	3	1	1
7	EditDate	1	1	3	3	1	1
8	EditTimes	1	1	3	3	1	1
9	FirstLineIndentAbsolute	1	1	3	3	1	1
10	Keywords	1	1	3	3	1	1
11	LeftMargin	1	1	3	3	1	1
12	LineHeightAutoValue	1	0	1	3	1	1
13	LineHeightFixedValue	1	1	3	3	1	1
14	PagesCount	1	1	3	1	1	1
15	ParagraphAfterAbsolute	1	1	3	3	0	1
16	ParagraphAfterRelate	1	0	1	3	0	1
17	ParagraphBeforeAbsolute	1	1	3	3	1	1
18	ParagraphBeforeRelate	1	0	1	3	1	1
19	ParagraphBorder	1	1	2	3	0	1
20	ParagraphsCount	1	1	3	3	1	1
21	RightMargin	1	1	3	3	1	1
22	RunBorder	1	0	1	1	1	0
23	Subject	1	1	3	3	1	1
24	SettingIndent	1	1	3	3	0	1
25	TextAlignment	1	1	2	3	1	0
26	TextBold	1	1	3	3	1	1
27	Title	1	1	3	3	1	1
28	TopMargin	1	1	3	3	1	1
29	WordsCount	1	1	3	1	1	1

参 考 文 献

- [1] ISO/IEC 26300 信息技术 面向办公应用的开放文档格式 (Information technology-Open Document Format for Office Applications)
- [2] ISO/IEC 29500 信息技术 文档描述与处理语言 开放办公文档XML格式 (Information technology-Document description and processing languages-Office Open XML File Formats)
-