



中华人民共和国国家标准

GB/T 27702—2011/ISO 23950:1998

信息与文献 信息检索(Z39.50) 应用服务定义和协议规范

Information and documentation—Information retrieval (Z39.50)—
Application service definition and protocol specification

(ISO 23950:1998, IDT)

2011-12-30 发布

2012-05-01 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

目 次

前言	I
引言	II
1 范围	1
2 规范性引用文件	1
3 术语和定义	2
4 信息检索服务	10
4.1 信息检索服务的模型和特征	11
4.2 信息检索服务机制	13
4.3 消息/记录长度和分段	53
4.4 操作和引用标识	58
4.5 并发操作	58
4.6 组成规范	59
4.7 Type-1 查询和 Type-101 查询	60
5 协议规范	63
5.1 Z39.50 APDU 的抽象语法和 ASN.1 规范	63
5.2 协议过程	79
5.3 扩展性规则	88
5.4 一致性	88
附录 A (规范性附录) OID:Z39.50 对象标识符	95
附录 B (规范性附录) CTX:应用环境 basic-Z39.50-ac	98
附录 C (规范性附录) ATR:属性集	99
附录 D (规范性附录) ERR:错误诊断	105
附录 E (规范性附录) REC:记录语法	114
附录 F (规范性附录) RSC:资源报告格式	137
附录 G (规范性附录) ACC:访问控制格式	139
附录 H (规范性附录) EXT:本标准定义的扩展服务	142
附录 I (规范性附录) USR:用户信息格式	157
附录 J (规范性附录) ESP:元素规范格式	160
附录 K (规范性附录) VAR:变量集	162
附录 L (规范性附录) TAG:标签集定义和模式	165
附录 M (资料性附录) ERS:扩展结构集模型	169
附录 N (资料性附录) RET:Z39.50 索取	171
附录 O (资料性附录) PRO:Z39.50 框架文件	189
附录 P (资料性附录) 维护机构的指定	190

前 言

本标准按照 GB/T 1.1—2009 给出的规则起草。

本标准等同采用国际标准 ISO 23950:1998《信息与文献 信息检索(Z39.50)应用服务定义和协议规范》。

本标准对国际标准 ISO 23950:1998 的中译文做了修改,技术内容未作变动。

本标准共有 16 个附录。

附录 A~附录 L 是规范性附录,附录 M~附录 P 是资料性附录。

本标准由全国信息与文献标准化技术委员会(SAC/TC 4)提出并归口。

本标准起草单位:中国国防科技信息中心、中国科学技术信息研究所、北京大学。

本标准主要起草人:真溱、龚昌明、汤珊红、李秀锦、赵晋巍、冯海涛、陈凌。

引 言

ISO 23950 与 ANSI/NISO Z39.50—1995 具有相同的内容,只是在文体上有一些细微的差别。例如,ANSI/NISO Z39.50 中的“附件”,ISO 23950 中称为“附录”。这两个标准具有相同的名称:“信息检索(Z39.50):应用服务定义和协议规范”。为避免认为这是两个不同的标准,在两个标准的名称中都明确地使用“Z39.50”。“Z39.50”通常指本标准定义的服务和协议。ISO 23950 被批准后,将取代相关标准 ISO 10162 和 ISO 10163。在本篇序言的其余部分,凡提到 Z39.50—1995 是指 ANSI/NISO Z39.50—1995,它与 ISO 23950 完全相同。凡提到 Z39.50—1988、Z39.50—1992 和 Z39.50—1994,是指与 ISO 23950 不同的较早的版本。

“ANSI/NISO Z39.50—1995 信息检索(Z39.50):应用服务定义和协议规范”是 ANSI/NISO Z39.50—1992 的修订版。Z39.50—1995 的草案稿是指 Z39.50—1994,实施者宜注意作为 Z39.50—1994 提出的所有草案都不是该标准的最新版。Z39.50—1994 的多种草案稿都是在 Z39.50—1995 之前提出的,Z39.50—1995 是最终定稿。

1984 年,本协议提出时仅用于书目信息领域。随着越来越多的专业领域对 Z39.50 产生兴趣,1990 年成立了 Z39.50 实施者小组(ZIG)。ZIG 的成员包括希望获取或提供各种信息的制造商、零售商、咨询商、信息提供商和大学,这些信息包括书目信息、文本信息、图像信息、金融信息、公共设施信息、化学信息和新闻。所有感兴趣的团体都可以成为 ZIG 的成员。

Z39.50 维护处成立于 1989 年,它隶属于美国国会图书馆。其任务是修订 Z39.50—1988,使之与国际标准 ISO 10162/10163 搜索与索取(SR)完全兼容。当时,为 1992 版提出了多个新的特性,以支持多种信息检索活动。但是,这些新特性尚未完全制订出来,如果把它们纳入 1992 版会大大地延误标准的公布。因此,维护处暂缓采纳新特性的提议,但同时向实施者承诺,将继续进行新特性的制订工作,下一版标准将成为与 1992 版兼容的超集。Z39.50—1992 取代了 Z39.50—1988,成为与 ISO 10162/10163 搜索与索取兼容的超集。

1992 年,维护处对 Z39.50 的实施者进行了一次正式调查,确定提议中的新特性的重要程度。调查的目的是:(a)把这些特性缩小到可管理的范围之内;(b)确定提出的新特性是否得到充分说明和理解;(c)估测它们的费用和复杂程度。调查结果表明有些特性是必不可少的,另有一些特性从更远的角度考虑可以删除。除此之外的其他特性,调查无法得出结论,对它们的处理最终根据多数人的意见决定。

Z39.50—1995 于 1991 年年底开始制定。从 1991 年 12 月到 1994 年 4 月,维护处在每次 ZIG 会议上都提出一个修订草案。实施者在 ZIG 的因特网邮件列表和会议上,仔细阅读和充分讨论每个草案。对每个草案的评论和意见,以及在 ZIG 会议上达成的协议都在后继的草案中得到体现。1994 年 4 月,ZIG 建议将草案最终定稿。

1992 版被认为是“第 2 版”,1995 版被认为是“第 3 版”。这些版本名确实具有特定的协议意义,但并不指标准的版本。Z39.50—1992 规定第 2 版协议;Z39.50—1995 规定第 2 版和第 3 版协议。

Z39.50—1992 代替了过时的 Z39.50—1988,但是,Z39.50—1992 和 Z39.50—1995 之间的关系则完全不同。Z39.50—1995 是与 1992 版兼容的超集,实施者可以从 Z39.50—1995 中得到第 2 版的全部细节,并实现与 Z39.50—1992 兼容。

协议基础

本协议规定了客户机和服务器之间信息交换的格式和过程,使客户机能够(a)请求服务器搜索某个数据库,并识别出满足特定要求的记录,(b)索取识别出的部分或全部记录。

客户机可以代表用户创建请求:本协议描述(可能位于不同的计算机上的)客户机和服务器之间的

通信;不描述客户机和用户之间的交互。

Z39.50—1992 和 ISO 23950 均支持下列基本功能:客户机可以发送搜索请求,指明一个或多个数据库,并包含一个查询以及用于确定搜索出的记录是否作为响应的一部分返回的参数。服务器以识别出的记录数作为响应,并可能返回部分或所有记录。然后,客户机可以索取选定的记录。客户机假设搜索到的记录形成一个“结果集”(一个有序集,其顺序由服务器确定),这些记录可以根据它们在结果集中的位置来引用。可选的功能包括:

某些情况下,客户机不希望接收完整的数据库记录时,可以指定一个数据元集,指明拟索取的数据元。例如,客户机可以指定“如果识别出的记录少于或等于 5 条,则传输‘完整’记录;若多于 5 条,则传输‘简要’记录”。

客户机可以为响应记录指定首选语法,例如:USMARC。

客户机可以命名结果集,以便以后引用。

客户机可以删除命名的结果集。

服务器可以在处理请求之前要求验证以便对客户机施加访问控制。

服务器可以通过发送非征求性或征求性的状态报告,提供资源控制;服务器可以暂停处理并允许客户机指明是否继续。

查询表达式

ISO 23950 对 Type-1 查询的支持做出详细的规定和要求。Type-1 查询用一个一个的搜索项表示,每个搜索项都有一个属性集,例如指明搜索项的类型(主题、名称等),是否被截断,以及搜索项的结构。服务器负责将这些属性映射到数据库的逻辑设计中。在一个 Type-1 查询中,搜索项可以通过布尔运算符组合起来。这些搜索项和运算符以逆波兰标记法表示。

属性集

与某个搜索项相关联的属性都属于某个特定的属性集,其定义是经过注册的,也就是被分配了一个唯一的、全局可识别的、包含在查询中的对象标识符,即属性集标识符。

附录 C ATR 定义并注册了属性集 bib-1,它描述了用于书目查询的各种属性。bib-1 属性集是由书目界开发的;如有需要,其他团体也可以开发并注册属性集。附加的属性集可以在 ISO 23950 之外进行注册。

响应记录

本协议将出现在服务器响应消息中的记录分为两类:数据库记录和诊断记录。

附录 E REC 注册了各种 MARC 格式的对象标识符,包括 USMARC、UKMARC、Norway MARC 和 CANMARC;服务器返回的数据库记录带有这些对象标识符。该附录定义了其他几种类型的记录格式,并规定了注册更多记录格式的方法。

诊断记录同样也带有标识其格式的对象标识符。附录 D ERR 定义并注册了两种诊断记录格式(Z39.50—1992 定义了其中一种),包括了用于书目应用的各种诊断代码。也可以注册其他的诊断记录格式。

新特性

下面汇总了 Z39.50—1995 的增强特性(相对于 1992 版)。“第 2 版”和“第 3 版”是协议的两个版本;“Z39.50—1992”和“Z39.50—1995”则是标准的两个版本。用“Z39.50—1995 的新特性”描述的某一特性一般适用于两个协议版本。例如扫描:实施者可以将扫描服务增加到已存在的 Z39.50—1992 实现中,而不必利用其他新特性。

下面分四类对增强特性进行描述:搜索、索取、新的服务和机制,以及其他增强的特性。

搜索

属性。属性和属性集具有许多新特性。在第 3 版中,单个查询(甚至单个搜索项)中的属性可以来自不同的属性集。这有两个优点:第一,这一特性可用于搜索多个数据库(虽然第 2 版支持对多个数据

库进行搜索,但一个查询包含的所有属性必须属于单个属性集,这就制约了搜索多个数据库的能力,除非这些数据库相似);第二,定义新的属性集时可以减少重复。

第3版增加了两个能够灵活地定义属性集的新特性。一是为属性值定义了新的数据类型(在第2版中,属性值只能是数值型)。二是属性集定义现在可以列出一些可选的评估规则(例如,是否允许服务器使用它认为更合适的属性来代替某个属性),并且查询时可选择其中一条规则。增强的 bib-1 属性集定义采用了这一新特性。

在 Z39.50—1995 中(除了包含 Z39.50—1992 中所有的属性),bib-1 定义还包含了许多新的属性。

扩展结果集模型。Z39.50—1992 中开发了结果集的基本模型;1995 版描述了“扩展结果集模型”,这一模型支持扩展的邻近搜索。

该扩展模型还支持第3版的一种新的搜索功能——限制,这项功能(实际上)是对结果集的一种操作。它允许根据指定的属性从结果集中选择记录。

搜索项。在第3版中,查询用的搜索项可以具有多种数据类型。(在第2版中,搜索项为二进制类型,实质上没有数据类型,通常用结构属性描述其类型。)由于减少了对结构属性的需要,该新特性将简化查询以及属性集定义。

中间结果。在 Z39.50—1995 中,服务器报告搜索进展时,可以提供每个查询部件(即每个子查询或每个数据库)的信息,作为搜索响应的一部分(仅用于第3版)或作为资源控制的一部分。服务器还可以为单个查询部件创建结果集并提供对结果集的访问。

索取

分段。在第2版中,一个索取响应被限制在单个消息中;服务器尽可能将请求的记录放入该消息中,如果放不下,则能放多少放多少。例如,客户机想要索取一万条记录,并且知道在单个消息中不可能索取这一万条记录。典型的作法是客户机请求所有的一万条记录,等待响应,确定索取到了多少条记录,然后为其余记录发送另一个请求。在许多环境下,这种方法很有效,但在高速网络中这样做会非常慢。服务器在发送每组记录前都必须等待请求,这就产生一个延迟。对于普通网络来说,这种延迟可以忽略不计,但对高速网络来说,这种延迟是不能忍受的。在第3版中,服务器可以用多个连续响应消息来回应一个索取请求,其间无需再穿插请求。

当单个记录太长、无法放入单个消息中时,出现更为严重的分段问题。第3版引入了二级分段:一个记录可以跨越多个响应消息。客户机或服务器可以选择支持任何一级的分段,或不分段(第2版规则适用这种情况)。

索取工具。为支持各种索取应用、特别是文献索取应用,ZIG 用了两年多时间开发出具有丰富索取功能的一个扩展模型及一套工具。附录 N 对该模型作了详细描述。Z39.50—1995 指定了几种新的对象类(模式、标签集和变量),并根据这些对象类和其他类定义了一些特定对象。该附录提供了这些对象的详细语义,描述了如何将它们结合起来使用提供多种文献索取功能。下面是几个例子:

一条数据库记录可以包括多篇文献。客户机可以找到并索取某篇指定文献,而不是整个数据库记录。

客户机可以索取某篇文献逻辑上或者物理上的某个特定部分,例如,特定的页、特定的章节、特定的标题、所有标题、或者所有的图像。客户机还可以只索取标题,例如所有章节或者所有节的标题。

可以用多种格式(如 PostScript、SGML)、语言、表示参数(如行长、每页行数和分栏等)以及其他变量获取某篇文献。客户机可以发现某篇文献支持哪些变量,以及与某种特殊变量形式相关的信息:例如,按照特定变量或者文献大小索取文献的费用。最终,客户机可以按照所需变量索取该文献(或某个特定部分)。

对于某个给定搜索,与某篇文献相关联的可能是命中点:它是指向(文献内)与搜索相关的搜索项的指针。客户机可以索取某篇文献中的命中点,以迅速定位到满意的那些部分。客户机也可以只索取(按重要性排序的)命中点,然后索取指定的满意的那些部分。

新的服务与机制

扫描和排序。扫描和排序是 Z39.50—1995 中新增的服务,前者用于对列表或索引中的搜索项进行扫描,后者用于对结果集进行排序。

扫描目前是 Z39.50 浏览机制中仅有的一种服务,但拟在今后的版本中增加其他浏览功能。

扩展服务群。扩展服务群是 Z39.50—1995 中的一种新机制,它包含一项新的 Z39.50 服务,即扩展服务群服务,用于创建 Z39.50 会话之外执行的一个特定的扩展服务任务,可以用 Z39.50 服务来监督该任务的进度。特定的扩展服务包括:保存结果集、建立周期性查询计划、输出文献、订购文献和更新数据库。

解释。新的解释机制允许客户机索取服务器实施情况的细节:一般特性(描述、联系信息、运行时间、限制、使用费用等)、可供搜索的数据库、索引、属性集、属性细节、模式、记录语法、排序能力和扩展服务。服务器在一个特定的数据库中维护解释信息,客户机可以利用 Z39.50 的搜索和索取机制来访问这个数据库。本标准详细描述了解释信息的格式。

有些解释信息对客户机是透明的,以便直接显示给客户机用户,并且也是这样指定的(例如“一般特性”)。有些解释信息可由客户机和用户共享。例如,客户机可以索取可搜索数据库的列表;客户机可以为列表中的每个数据库显示一个非正式的名称、一个图标,和一段简要的描述。同时,客户机可以在协议消息中保留使用的真实的数据库名称,并可以不显示这个真名。有些解释信息对用户是完全透明的。例如,客户机可以索取某个数据库支持的属性信息,并在形成一个查询表达式时(即把用户提供的查询转化为一个 Z39.50 的 Type-1 查询时),使用该信息。

其他增强的特性

终止和重新初始化。第 3 版包括更为灵活地终止 Z39.50 会话的方法,这实际上允许在不中断网络连接的情况下重新初始化。

并发操作。第 3 版允许多个并发操作。而在第 2 版中,操作是严格按顺序执行的。

诊断。大多数 Z39.50 服务都包括诊断功能。在第 2 版中,一个诊断必须与本标准定义的特定格式一致。在第 3 版中,诊断格式可以在外部定义和注册。本标准定义了这样一种(新的)诊断格式,附带了一个综合的诊断集。

访问控制格式。Z39.50—1992 提供了访问控制,但未定义任何访问控制格式。Z39.50—1995 定义了加密和认证格式,并定义了允许服务器向客户机提示各种信息的格式。

字符集支持。为字符串引入了一种新的数据类型,即“International String”(国际字符串)。它的定义使客户机和服务器能更灵活地决定在一次会话中使用某一特殊语言以及一种或多种字符集。

计量单位。为支持计量单位,引入了新的数据类型。这些定义允许使用标准的方法表示计量单位的类型及计量单位。例如,计量单位的类型是“质量”,计量单位是“千克”。

可扩展性和协商。第 3 版提供了强大的可扩展性。每个协议消息都包含一个字段,用于标示在外部定义格式的那些信息。这些外部定义的格式作为 ISO 23950 的临时性扩展和试验性使用,由 Z39.50 维护处注册和管理,并有可能并入将来的版本。

Z39.50—1995 引入了“协商记录”的概念。客户机在初始化消息中可以包含协商记录,以建议某些条件是否在会话中起作用(例如,使用某一特殊语言以及一种或多种字符集)。服务器可以响应,指明建议是否被接受或提出一个反建议。

协商记录是新的可扩展特性的应用。协商记录可以在外部定义,并由 Z39.50 维护处维护。

信息与文献 信息检索(Z39.50)

应用服务定义和协议规范

1 范围

本标准规定了信息检索的应用服务定义和协议规范。服务定义部分描述了一个应用中实现搜索和索取数据库中信息的服务；协议规范部分包括协议控制信息的定义、交换该信息的规则，以及实施协议需要满足的一致性要求。

本标准适用于支持信息检索服务的系统，以及诸如信息服务机构、大学、图书馆、联合编目中心这样的组织。

注：Z39.50 共有三种出版物：Z39.50—1988、Z39.50—1992 和 Z39.50—1995；搜索和检索协议有一种出版物：ISO 10163-1:1993、Z39.50—1992、ISO 10163-1:1993 和 Z39.50—1995（不包括 Z39.50—1988）这三种出版物都体现了协议版本概念，并定义了三种协议版本：第 1 版、第 2 版和第 3 版。ISO 10163-1:1993 基于第 1 版协议；Z39.50—1992 基于第 2 版协议；Z39.50—1995 基于第 2 版协议和第 3 版协议。（Z39.50—1988 与协议版本无关。）本标准基于第 2 版和第 3 版协议。它假定第 1 版和第 2 版是一致的，支持第 2 版的实施自动支持第 1 版（另外，在本标准的其他任何地方都不直接提及第 1 版）。对本标准中只适用第 2 版或第 3 版的过程，分别给出了注释。

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 15695—1995 信息技术 开放系统互连 表示服务定义(ISO/IEC 8822:1994, IDT)

GB/T 16262—1996(所有部分) 信息技术 抽象语法记法(ASN.1)规范(ISO/IEC 8824:2002, IDT)

GB/T 16263—1996(所有部分) 信息技术 ASN.1 编码规则(ISO 8825:2002, IDT)

GB/T 16687.1—2008 信息技术 开放系统互连 面向连接的联系控制服务元素协议 第 1 部分:协议规范(ISO 8650-1:1996, IDT)

GB/T 16688—2008 信息技术 开放系统互连 联系控制服务元素服务定义(ISO/IEC 8649:1996, IDT)

ANSI/NISO Z39.53—1994 信息交换语言表示代码(Codes for the representation of languages for information interchange)

ANSI/NISO Z39.58—1992 联机交互信息检索的公共命令语言(Common command language for online interactive information retrieval)

ISO 2709:1996 信息和文献 信息交换格式(Information and documentation—Format for information exchange)

ISO 4217:1990 货币和现金表示代码(Codes for the representation of currencies and funds)

ISO 7498:1984 信息处理系统 开放系统互连 基本参考模型(Information processing systems—Open systems interconnection—Basic reference model)

ISO 8649:1987 信息处理系统 开放系统互连 联系控制服务元素服务定义(Information pro-