



[12] 发明专利申请公开说明书

[21] 申请号 97118526.3

[43]公开日 1998年4月22日

[11] 公开号 CN 1179576A

[22]申请日 97.9.12

[30]优先权

[32]96.10.15[33]JP[31]272755 / 96

[71]申请人 国际商业机器公司

地址 美国纽约

[72]发明人 小早川健 田添英一 石仓纯

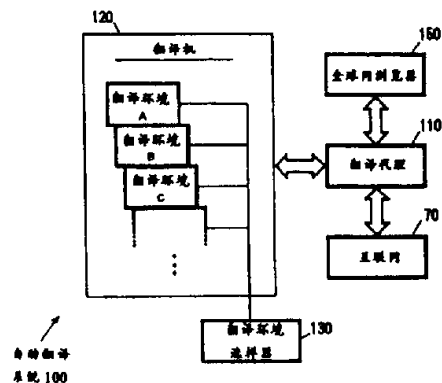
[74]专利代理机构 中国国际贸易促进委员会专利商标
事务所
代理人 王以平

权利要求书 6 页 说明书 22 页 附图页数 7 页

[54]发明名称 自动翻译方法和自动翻译机

[57]摘要

具有多个翻译环境的一种高级自动翻译系统，根据本发明的计算机处理过程包括将外语文本翻译成本地语言文本的翻译机；多个翻译环境；以及关于 URL 字符串和翻译环境的数据库。当用户在全球网浏览器屏上输入某个 URL 字符串时，检查该数据库以查找最接近该输入的 URL 字符串的某个 URL 字符串，翻译机利用对应于所检索到的 URL 字符串的翻译环境来翻译从所输入的 URL 字符串指定的全球网页上下载的文件中的文本。





权 利 要 求 书

1. 由装有多个翻译环境的翻译系统执行的一种自动翻译方法，用于翻译存放在由用户指定的某个信息资源中的文件中的文本，其特征在于包括步骤：

(a) 解释某个组合的字符串，在该字符串中，访问信息源的装置和所述信息源的名称被指定；

(b) 根据解释的结果，选择所述多个翻译环境中的一个；

(c) 使用所选择的翻译环境，翻译存放在由所述的组合字符串指定的所述信息源中的文件中的文本。

2. 由装有多种翻译环境的翻译系统执行的一种自动翻译方法，用于翻译存放在由用户指定的信息源中的某个文件中的文本，其特征在于包括步骤：

(a) 用户输入某个组合的字符串，在该字符串中，访问信息源的装置和所述信息源的名称被指定；

(b) 根据所述组合的字符串所指定的访问装置，下载存放在由所述的组合字符串指定的所述信息源中的某个文件；

(c) 解释所述的组合字符串；

(d) 根据解释的结果，选择所述的多个翻译环境中的一个；并且

(e) 使用所选择的翻译环境，翻译被下载文件中的文本。

3. 由装有多种翻译环境的翻译系统执行的一种自动翻译方法，用于翻译存放在由用户指定的信息源中的某个文件中的文本，其特征在于包括步骤：

(a) 用户输入某个组合字符串，在该字符串中指定信息源的访问装置和所述信息源的名称；

(b) 根据所述组合字符串指定的访问装置，下载存放在由所述组合字符串指定的所述信息源中的某个文件

(c) 在显示屏上显示下载的文件以及所述文件中的原始文件；

(d) 解释所述的组合字符串；

(e) 根据解释结果，从所述的多个翻译环境中选择一个；



(f)使用所选择的翻译环境翻译所述的原始文件;并且

(g)在显示屏上显示被翻译的文本以及原始文本。

4.由装有多个翻译环境的翻译系统执行的一种翻译方法,用于下载存放在由说明了所述信息源访问装置和所述的信息源的名称的某个组合字符串指出的信息源中的某个文件并翻译所述文件中的文本,其特征在于包括步骤:

(a)作为一个数据库,记录所述组合字符串和与被指定的信息源相关的某个翻译环境的一种组合;

(b)在所述的数据库中查找最类似某个新输入的组合字符串的组合字符串;并且

(c)使用对应于在所述的步骤(b)中找到的组合字符串的翻译环境,翻译下载文件中的文本。

5.由装有多个翻译环境的翻译系统执行的一种自动翻译方法,下载存放在由说明了所述信息源以及所述信息源的名称的用户输入组合字符串指定的信息源中的某个文件并翻译所述文件中的文本,其特征在于包括步骤:

(a)作为一个数据库,记录所述组合字符串和与被指定的信息源相关的翻译环境的一种组合;

(b)在所述的数据库中查找某个组合的字符串,该字符串最类似某个新输入的组合字符串;

(c)利用对应于在所述的步骤(b)中找到的组合字符串的翻译环境,翻译从被指定的信息源中下载的文件中的文本;并且

(d)在没有得到适当的翻译结果的情况下,使用另一个翻译环境来重新翻译该文本。

6.由装有多个翻译环境的翻译系统执行的一种自动翻译方法,下载存放在由说明了所述信息源的访问装置和所述信息源的名称的用户输入的组合字符串指定的信息源中的某个文件并翻译所述文件中的文本,其特征在于包括步骤:

(a)作为数据库,记录下所述组合字符串和与被指定的信息源有关的翻译环境的一种组合;



(b)在所述的数据库中查找某个组合的字符串,该字符串最类似某个新输入的组合字符串;

(c)使用对应于在所述的步骤(b)中找到的组合字符串的翻译环境,翻译从被指定的信息源中下载的文件中的文本;

(d)在没有得到适当的翻译结果的情况下,使用另一个翻译环境重新翻译所述的文本;并且

(e)在得到正确的翻译结果的情况下,在所述的数据库中另外再记录新的组合字符串和被使用的翻译环境之间的关系。

7.一种自动翻译方法,该方法根据某个由访问信息源的协议名和存放所述信息源的服务器名(以及在所述服务器中的某个文件名)组合的某个URL(均匀资源定位)字符串,通过互联网下载被指定服务器上的某个文件,并利用多个翻译环境中的一个翻译被下载文件中的文本,其特征在于包括步骤:

(a)解释该URL字符串;

(b)根据解释结果从所述的多个翻译环境中选择一个;并且

(c)利用所选择的翻译环境来翻译下载文件中的文本。

8.一种自动翻译方法,该方法根据某个访问信息源的协议名和存放所述信息源的服务器名(以及在所述服务器中的某个文件名)组合而成的URL(均匀资源定位)字符串,通过互联网下载在指定服务器上的文件,并利用多个翻译环境中的一个翻译被下载文件中的文本,其特征在于包括步骤:

(a)在显示屏上显示被下载的文件及原始文本;

(b)解释该URL字符串;

(c)根据解释结果从所述的多个翻译环境中选择一个;

(d)利用所选择的翻译环境来翻译下载文件中的文本;并且

(e)在所述显示屏上显示翻译文本和原文。

9.一种自动翻译方法,该方法根据由某个访问信息源的协议名称和存放所述信息源的服务器名(以及在所述的服务器中的文件名)组合而成的URL(均匀资源定位)字符串,通过互联网下载被指定服务器上的文件,并利用多个翻译环境中的一个来翻译被下载文件中



的文本,其特征在于包括步骤:

- (a)作为数据库记录 URL 字符串和翻译环境的一种组合形式;
- (b)在所述的数据库中查找某个 URL 字符串,该字符串最类似某个新输入的 URL 字符串;并且
- (c)利用对应所述的步骤(b)中得到的 URL 字符串的翻译环境,翻译被下载文件中的文本。

10.一种自动翻译方法,该方法根据由访问信息源的协议名称和存放所述信息源的服务器名组合而成的某个 URL(均匀资源定位)字符串,通过互联网下载被指定服务器上的文件,并利用多个翻译环境中的一个来翻译被下载文件中的文本,其特征在于包括步骤:

- (a)作为数据库,记录所述 URL 字符串和某个翻译环境的某种组合形式;
- (b)在所述的数据库中查找某个 URL 字符串,该字符串最类似某个新输入的 URL 字符串;
- (c)利用对应所述的步骤(b)中得到的 URL 字符串的某个翻译环境,翻译下载文件中的文本;并且
- (d)在得不到正确的翻译结果的情况下,使用另一种翻译环境重新翻译该文本。

11.一种自动翻译方法,该方法根据由访问信息源的协议名和存放所述信息源的服务器的名称(以及在所述服务器中的文件名)组成的 URL(均匀资源定位)字符串,通过互联网下载指定服务器中的文件,并利用多个翻译环境中的一个来翻译下载文件中的文本,其特征在于包括步骤:

- (a)作为数据库,记录所述 URL 字符串和某个翻译环境的某种组合形式;
- (b)在所述的数据库中查找某个 URL 字符串,该字符串最接近某个新输入的 URL 字符串;
- (c)使用和所述的步骤(b)中得到的 URL 字符串对应的某个翻译环境,翻译下载文件中的文本;
- (d)在得不到正确的翻译结果的情况下,利用另一个翻译环境重



新翻译所述的文本;并且

(e)在得到正确的翻译结果的情况下,另外再记录新的 URL 字符串和该翻译环境之间的关系。

12.一种自动翻译机,使用多个翻译环境中的一个来翻译借助互联网从某个 WWW(全球网)服务器中下载的文件中的文本,其特征在于包括:

(a)存放数据库的第一装置,该数据库管理 URL 字符串和翻译环境的组合形式;

(b)输入某个 URL 字符串的第二装置;

(c)从某个 WWW 服务器中下载由所述的输入 URL 字符串指定的文件的第三装置;

(d)在所述的第一装置中查找某个 URL 字符串的第四装置,该字符串最接近新输入的 URL 字符串;

(e)选择和被查找的 URL 字符串相关的某个合适的翻译环境的第五装置;

(f)利用所选择的翻译环境,翻译从 WWW 服务器中下载的文件中的文本的第六装置;以及

(g)在显示屏上显示翻译结果的第七装置。

13.一种计算机可编程存储介质,存放某个计算机程序,用于翻译从互联网的某个 WWW 服务器中下载的文件中的文本,所述的计算机程序的特征在于包括:

(a)多个翻译环境;

(b)记录某个表的数据库,在该表中,管理着 URL 字符串和所述翻译环境的组合;

(c)第一计算机程序代码装置,下载由用户以 URL 字符串的形式指定的 WWW 服务器上的文件;

(d)第二计算机程序代码装置,在所述的数据库中,查找某个 URL 字符串,该字符串最接近新输入的某个 URL 字符串;

(e)第三计算机程序代码装置,选择与被查找出来的 URL 字符串相关的某个适当的翻译环境;



(f) 第四计算机程序代码装置, 利用所选择的翻译环境, 翻译由 WWW 服务器上下载的文件中的文本; 以及

(g) 第五计算机程序代码装置, 在显示屏上显示该文本的翻译结果。



说明书

自动翻译方法和自动翻译机

本发明涉及一种自动翻译方法和自动翻译机,将外语形式(例如英文)的文本翻译为某种本地语言形式(例如日文或朝鲜文、中文等等)的文本,尤其涉及装有多种翻译环境的自动翻译方法和自动翻译机。更具体地说,本发明涉及一种自动翻译方法和自动翻译机,将由互联网上的某个 WWW 服务器提供的 WWW 页中的外语文本翻译为本地文本,尤其涉及这样的一种自动翻译方法和机器,能自动地选择一种适合于该外语文本的翻译环境来翻译外语文本。在这种情况下,“翻译环境”意味着用于翻译的某个字典数据库和文法算法(例如,子句的设置,辅助动词定义的设置,风格的选择)。

互联网:

利用类似公共电话线这样的通信线路远程地连接本地化计算机的“远程计算系统”已经得到研究和开发。例如,在日本,1985年电信工业法开始生效后,对利用个人计算机的通信普遍感兴趣。

最新的主要远程计算活动涉及到互联网的使用。互联网是由多个计算机网络协同互连而形成的一个网络,是70年代在美国的国家科学基金会(NSF)投资建立的网络“NSFnet”的基础上形成的。互联网作为一个国际网络,最初连接54个国家的大学和研究所,用于交换学术研究信息。然而,安装在大学和研究所的服务器(主要是UNIX工作站)随意地继续互相连接,使得互联网已经成为一个庞大的世界范围的网络。目前的互联网是由世界范围内的超过几十万个服务器构成的。利用所谓的“TCP/IP(传输控制协议/互联网协议)”通信协议将远程定位的服务器互相连接起来,并且通过传送数据或借助BBS(电子公告牌系统)可以互相通信。

传统的(例如最初的)互联网或多或少被看作基于计算机连接的一种UNIX系统。然而,随着1994年对公众的开放以及连接服务提供商的出现,互联网已经逐渐地被普通办公室和个人所使用。例



如,在日本,主要的个人计算机通信服务(例如“PC - VAN”和“NiftyServe”)提供互联网的连接服务。

互联网是由不同组织运行的并且具有不同策略的各种网络的聚集。尽管各个组织管理着各自的网络,但没有一个组织对互联网提供全局的管理。即,世界范围的各种服务器,例如数据库服务器和文件服务器,由网络的各自的用户局部管理。互联网的用户也可以通过某个服务器连接互联网而成为信息的提供者,并且可以严密地存取互联网上任何类型的信息。这和个人计算机服务有很大的区别,通过个人计算机服务,用户只能访问某个特定的主机,即某个信息服务器。在英文中,“互联网”通常被看作一个合适的名词,用以区分“互联网”和为特定公司的计算机互相连接而建立起来的广域网。

全球网(WWW):

毫无疑问,“全球网(WWW)”的引入加速了互联网的普及进程。全球网是一个广域信息系统(互联网应用),在网络(尤其是在互联网)上建立超文本^{/*}以便严密地存取信息,全球网最初是由欧洲粒子物理实验室(CERN)在1989年发展起来的。利用全球网,用户,即全球网客户,可以从服务器跳到另一个服务器来检索目标信息,或者只需用鼠标选择屏幕上醒目的字、词组或图像,就能从数据库中检索所要的信息。

根据“全球网”这一名称的含义,目前,互联网上有大量的全球网服务器。由称为HTML(超文本标记语言)的语言描述的多个文件构成的多个全球网页累积在每个全球网服务器的本地磁盘上。

包含在HTML文件中的是文本文件和其它各种类型的文件,例如图像、语音和图像混合文件。HTML通常基于SGML(标准通用化标记语言),并且,其当前的版本“HTML 3.X”是由全球网标准化组织“W3 Consortium”在1996年发行的。符合HTML的文件基本上是由文本文件构成的,并且文本的格式、图像文件和语音文件的定位以及链接目标都是用“标记”描述的,而“标记”则由辅助文件分类的“<”和“>”来定界。

全球网用户(全球网客户)可以借助称为HTTP(超文本传输协



议)(HTTP 对应 TCP/IP 的上层协议)的通信协议读出存放在全球网服务器的全球网页中的文件。所谓的“全球网浏览器”,一种用作互联网导游的计算机程序,能解释 HTML 文件的内容并将这些内容显示在用户机器的屏幕上。全球网浏览器的典型例子为 IBM 公司生产的“Web Explorer”(“Web Explorer”为 IBM 公司的商标)、NetScape 通信公司生产的“Netscape Navigator”以及微软公司生产的“Internet Explorer”。

HTML 文件中设计的图像文件可以在不同的全球网服务器上被定位。全球网浏览器不但能解释 HTML 文件,而且能读图像文件,并且能将包含有图像的组合文本显示在全球网客户机的显示屏上。此外,HTML 文件包含类似“当该文本(或图像)被鼠标选中时,读哪个服务器的什么文件”这样的命令。用户只需用鼠标选择出现在显示屏上的文本或者点屏幕上的某个位置,就能读出另一个 HTML 文件。由鼠标选择的 HTML 文件或者定位在与其前面的 HTML 文件相同的全球网服务器上,或者定位在另一个全球网服务器上。

将指示读下一个 HTML 文件的命令包含在 HTML 文件中,在该技术领域中也称为“形成一个链”。链接目标是用称为 URL(均匀资源定位)的标识符指定的。URL 是用于识别互联网上的信息资源(即由服务器存放在互联网上的文件)的一种标准的寻址技术。URL 的规格由 RFC1738 规定,这是由 IETF(互联网工程任务组织)/**/准备的一个正式的 RFC(要求注释)文档集。

组成 URL 的字符串包括三个部分:访问信息源的协议名(模式名);存放信息源的服务器名;以及该服务器中的文件名。可选择的协议名为 HTTP(http)、NNTP(news)、FTP(ftp:文件传输协议)、Telnet(Telenet)以及访问本地磁盘(文件)。典型的服务器名为“www.ibm.com”、“www.yahoo.com”等等。由服务器“www.yahoo.com”的全球网(WWW)服务提供的文件“Arts”的 URL 被表示为“http://www.yahoo.com/Arts/”。URL 的各部分之间用“/”隔开。换句话说,URL 是全球网页的地址。应该注意:“全球网页”是



由某个全球网服务器提供的屏幕的一般名称,也称为“主页(home page)”。

在构成 URL 的字符串中,根据该规格说明,由“/”隔开的每个字符串段以下称为“部分 URL 字符串”。

全球网页的翻译程序:

如上所述,使用全球网,互联网的用户坐在桌子前能访问世界范围内的全球网服务器,并且在浏览器屏幕上可以看到所需的全球网页。

由于全球网页中的大部分文本是用英文写的,因此,使用英语区外面的互联网用户(例如在日本、中国或朝鲜)或者不熟悉英文的其他互联网用户读这样的全球网页都会麻烦,因此就无法充分利用这种广域信息系统。因此,自动翻译软件或自动翻译支持软件最近已经得到开发和生产。某些自动翻译软件程序是专门为互联网上的文本(即全球网页)的翻译而设计的。自动翻译软件利用全球网浏览器的代理功能(Proxy function)/* */转达经互联网得到的数据,被翻译的文档重新被传送到全球网浏览器并在浏览器屏幕上显示。由承诺和互联网兼容而产生的自动翻译软件的一个例子是由 IBM 日本公司销售的“King of Internet Translation” 1.X 版。

自动翻译是这样一种技术,通过这种技术,计算机系统的处理能力被用来将用某种外国语言(例如英语)准备的文本翻译成为用本地语言(例如日语)表达的文本,或反之。按照人类处理相关语言的庞大知识库的模型建立数据库,并且当执行自动翻译时,翻译机引用该数据库。

准确地实现对原始文本的自动翻译而将原始文本的内容完全忠实地反映在翻译中,这样做通常是很困难的。这是由于自然语言本身固有的二义性所造成的直接结果。这种二义性的一个根源是由于所用的词汇具有多种意义,即具有两种或多种意义,而这经常出现在原始文本中。每当具有多种意义的一个词出现在文本中时,翻译机必须选择一种意义,而这种意义的选择伴随着某些错误,即错译。二义性的另一种来源是语法变化。例如,英文句子被划分为几种基本

句型,例如 SVO, SVOO 和 SVOC,这是从主语、动词等之间的关系来划分的。当省略关联词(“that”, “which”, “why”等)时,理解句型和语法是很困难的。

提供多种翻译环境的方法已经得到广泛的使用,以保证更加准确的自动翻译。术语“翻译环境”描述了用于某个翻译过程中的字典数据库和文法算法(子句设置和辅助动词意义设置以及句子风格设计),并且一般认为原文所属的风格、领域和应用必须和某个翻译环境相关。同样的词用在不同的领域(例如艺术、运动、教育和科学)时,通常具有不同的意义和用法,而且用在不同方面的特定翻译环境,例如互联网字典、艺术字典和运动字典,都已经被用在实际中。当使用适合原文内容的翻译环境时,可以去掉自动翻译期间的不必要的分析,并且提高翻译精度。

选择适当翻译环境最容易和基本的方法是根据原文的内容,由用户手工选择合适的翻译环境。在这种情况下,执行自动翻译之前,用户必须理解原文的内容。然而,用户很难对用一种非本地语言的语言写的原文的内容一看就能理解,而用户可能要花许多时间才能理解文本,结果是,翻译系统的提供就没有太大的意义。尤其是因为许多互联网用户当选择有用的信息时需要自动翻译功能的帮助,以便扫描大量的全球网页,如果看原文需要许多时间,自动翻译将是一种毫无意义的活动。

还有另一种背景技术,也用在自动翻译系统中;负责选择某个翻译环境。例如,在日本未审查的专利发布号 Hei7-191999, Hei6-332946, Hei6-318229, Hei6-60117 和 Sho 61-173060 中,提出了一种自动选择翻译环境的翻译系统。这种翻译系统消除了用户选择某个翻译环境时所要求的麻烦过程,因此提高了工作效率。然而,在选择匹配原文内容的一种翻译环境之前,大多数常规技术分析原文的内容,分析原文的过程需要相当长的时间才能完成,这就使得翻译系统变得复杂。换句话说,从其结构和其使用两方面来看,常规技术都是复杂的系统。

《注释》



* :超文本是一种软件,将字符、图形和表格作为目标并提供目标之间的各种相互关系,以便访问这些目标。

* * :IETF 是一个组织,研究互联网上的各种技术问题,讨论服务并提供解决方案。

* * * :代理功能是这样一种功能,当主机服务作为代理时,接收来自互联网的通信数据,并将这些数据转送给不会影响直接外部访问的另一个主机。

本发明的一个目的是提供一种优良的自动翻译方法和机器,利用装有多个翻译环境的翻译系统,将外文文本(如英文文本)翻译成本国语言文本(如日文、朝鲜文或中文文本)。本发明的另一个目的是提供一种优良的自动翻译方法和机器,在将外语文本翻成本国语言文本之前,自动选择适合于将被翻译的外语文本的一种翻译环境。

本发明的又一个目的是提供一种优良的自动翻译方法和机器,利用装有多种翻译环境的翻译系统,以便将互联网上的全球网服务器提供的全球网页中的外语文本适当地翻成本地语言的文本。

为了实现上述的目标,根据本发明的第一方面,由装有多个翻译环境的翻译系统执行的一种自动翻译方法,翻译存放在由用户指定的信息源中的某个文件中的文本,包括步骤:(a)解释某个组合的字符串,在该字符串中,访问信息源的装置和信息源的名称被指定;(b)根据解释得到的结果选择多个翻译环境中的一个;以及(c)使用所选择的翻译环境,翻译存在由组合的字符串指定的信息源中的文件中的文本。

根据本发明的第二方面,由装有多个翻译环境的翻译系统执行的一种自动翻译方法,翻译存放在由用户指定的信息源中的某个文件中的文本,包括步骤:(a)用户输入某个组合的字符串,在该字符串中,信息源的访问装置以及信息源的名称被指定;(b)根据组合字符串指定的访问装置,下载存放在由组合字符串指定的信息源中的某个文件;(c)解释该组合字符串;(d)根据解释的结果选择多个翻译环境中的一个;并且(e)利用所选择的翻译环境翻译下载文件中的文



本。

根据本发明的第三方面,由装有多个翻译环境的翻译系统执行的一种自动翻译方法,翻译存放在由用户指定的信息源中的某个文件中的文本,包括步骤:(a)用户输入某个组合字符串,在该字符串中,对信息源的访问装置和信息源的名称被指定;(b)根据组合字符串中所指定的访问装置,下载存放在由组合字符串指定的信息源中的文件;(c)在显示屏上显示被下载的文件以及该文件中的原始文件;(d)解释该组合字符串;(e)根据所解释的结果,选择多个翻译环境中的一个;(f)利用所选择的翻译环境翻译原文;并且(g)在显示屏上显示被翻译的文本和原文。

根据本发明的第四方面,由装有多个翻译环境的翻译系统执行的一种自动翻译方法,下载存放在由用户输入的并指定了对信息源的访问装置和信息源的名称的组合字符串指定的信息源中的文件,并翻译该文件中的文本,包括步骤:(a)作为数据库记录该组合字符串和与所指定的信息源相关的某个翻译环境的组合形式;(b)在该数据库中查找最接近新输入的某个组合字符串的某个组合字符串;并且(c)利用对应于在步(b)中找到的组合字符串的翻译环境翻译被下载文件的文本。

根据本发明的第五方面,由装有多个翻译环境的翻译系统执行的一种自动翻译方法,下载存放在由用户输入的并指定了对信息源的访问装置和信息源的名称的组合字符串指定的信息源中的某个文件,并翻译该文件中的文本,包括步骤:(a)作为数据库记录该组合字符串和与被指定的信息源相关的某个翻译环境的某种组合形式;(b)在该数据库中查找某个组合字符串,该字符串最接近新输入的某个组合字符串;(c)利用对应于在步(b)中找到的组合字符串的翻译环境,翻译从被指定的信息源中下载的文件中的文本;并且(d)在没有获得正确翻译结果的情况下,利用另一个翻译环境重新翻译该文本。

根据本发明的第六方面,由装有多个翻译环境的翻译系统执行的一种自动翻译方法,下载存放在由用户输入的并指定了对信息源的访问装置和信息源的名称的某个组合字符串指定的信息源中的某



个文件,并翻译该文件中的文件,包括步骤:(a)作为数据库记录该组合字符串和与被指定信息源相关的某个翻译环境的某种组合形式;(b)在该数据库中,查找某个组合字符串,该字符串最接近新输入的某个组合字符串;(c)利用对在步(b)中找到的组合字符串的翻译环境,翻译从被指定的信息源中下载的文件中的文本;(d)在得不到正确翻译结果的情况下,利用另一个翻译环境重新翻译该文本;并且(e)在得到正确翻译结果的情况下,在数据库中另外再记录新的组合字符串和所使用的翻译环境之间的关系。

根据本发明的第七方面,一种自动翻译方法,根据作为访问某个信息源的协议名和存放信息源的服务器的名称(以及服务器中的某个文件名)的组的 URL(均匀资源定位)字符串,通过互联网下载在某个被指定的服务器中的文件,并利用多个翻译环境中的一个,翻译被下载文件中的文本,包括步骤:(a)解释该 URL 字符串;(b)根据某个解释结果选择多个翻译环境中的一个;并且(c)利用所选择的翻译环境翻译被下载文件中的文本。

根据本发明的第八方面,一种自动翻译方法,根据作为访问某个信息源的协议名和存放该信息源的服务器的名称(以及该服务器中的某个文件名)的一种组合形式的 URL(均匀资源定位)字符串,通过互联网下载某个被指定的服务器中的文件,并利用多个翻译环境中的一个翻译被下载文件中的文本,包括步骤:(a)在显示屏上显示被下载文件以及原始文本;(b)解释该 URL 字符串;(c)根据某个解释结果,选择多个翻译环境中的一个;(d)利用被选择的翻译环境翻译原始文本;并且(e)在显示屏上显示翻译的文本和原始文本。

根据本发明的第九个方面,一种自动翻译方法,根据作为访问某个信息源的协议名和存放信息源的服务器的名称(以及该服务器中的某个文件名)的一种组合形式的 URL(均匀资源定位)字符串,通过互联网下载所指定服务器中的文件,并利用多个翻译环境中的一个翻译被下载文件中的文本,包括步骤:(a)作为数据库记录 URL 字符串和某个翻译环境的组合;(b)在该数据库中查找某个 URL 字符串,该字符串最接近新输入的某个 URL 字符串;并且(c)利用对在



步(b)中找到的 URL 字符串的某个翻译环境,翻译被下载文件中的文本。

根据本发明的第十个方面,一种自动翻译方法,根据作为访问某个信息源的协议名和存放该信息源的服务器的名称(以及该服务器中的某个文件名)的一种组合形式的 URL(均匀资源定位)字符串,通过互联网下载被指定服务器中的文件,并利用多个翻译环境中的一个翻译被下载文件中的文本,包括步骤:(a)作为数据库记录该 URL 字符串和某个翻译环境的某种组合形式;(b)在该数据库中查找某个 URL 字符串,该字符串最接近新输入的某个 URL 字符串;(c)利用对应于在步(b)中得到的 URL 字符串的翻译环境,翻译被下载文件中的文本;并且(d)在得不到正确翻译结果的情况下,利用另一种翻译环境重新翻译该文本。

根据本发明的第十一方面,一种自动翻译方法,根据作为访问某个信息源的协议名和存放该信息源的服务器的名称(以及该服务器中的某个文件名)的某种组合形式的 URL(均匀资源定位)字符串,通过互联网下载被指定服务器中的文件,并利用多个翻译环境中的一个,翻译被下载文件中的文本,包括步骤:(a)作为数据库记录该 URL 字符串和某个翻译环境的一种组合形式;(b)在该数据库中查找某个 URL 字符串,该字符串最接近新输入的某个 URL 字符串;(c)利用对应于在步(b)中得到的 URL 字符串的翻译环境,翻译被下载文件中的文本;(d)在没有得到正确翻译结果的情况下,利用另一个翻译环境重新翻译该文本;并且(e)在得到正确的翻译结果的情况下,将新的 URL 字符串和该翻译环境之间的关系也记录在数据库中。

根据本发明的第十二方面,一种自动翻译机器,使用多个翻译环境中的一个翻译通过互联网从某个全球网服务器中下载的某个文件中的文本,包括:(a)存放某个数据库的第一装置,该数据库管理 URL 字符串和某个翻译环境的组合形式;(b)输入 URL 字符串的第二装置;(c)从某个全球网服务器中下载由输入的 URL 串指定的文件的第三装置;(d)在第一装置中查找某个 URL 字符串的第四装



置,该字符串最接近新输入的某个 URL 字符串;(e)选择和被查找到的 URL 字符串相关的翻译环境作为合适的一个翻译环境的第五装置;(f)利用所选择的翻译环境,翻译从全球网服务器中下载的文件中的文本的第六装置;并且(g)将翻译结果显示在屏幕上的第七装置。

“翻译环境”为某个字典数据库和文法算法(例如,设置子句,设置辅助动词的意义,指定句子的风格),用在翻译处理过程中。

本发明主要用于全球网页的自动翻译,并且基于这样的事实,作为某个全球网页的地址的 URL 字符串由描述性字符串段的组合构成(即“部分 URL 字符串”):协议名,服务器名和文件名。

根据本发明的自动翻译方法和机器可以借助计算机处理的方式来实现,通过和从全球网服务器中接收全球网页并将其在浏览屏上显示的一系列浏览操作互锁,实现这种计算机处理。

这种计算机处理包括翻译机,用于将外国语言的文本翻译为本地语言的文本(或反之),还包括多种翻译环境,例如翻译环境 A(互联网字典)、翻译环境 B(艺术字典)以及翻译环境 C(运动字典)。根据本发明的自动翻译方法,用于 URL 字符串的数据库被累积。该数据库可以采用表的格式,如表 1 所示,该表格给出了与相当频繁访问的全球网页相关的 URL 字符串和翻译环境之间的对应关系。根据该翻译过程的过去执行情况得到的结果,对应各个 URL 字符串的翻译环境被认为是适当的环境。

[表 1]

在表 1 中,翻译环境 A(互联网字典)被记录为 URL 名“www.ibm.com/”和“www.yahoo.com/”的适当环境;翻译环境 B(艺术字典)被记录为“www.yahoo.com/Arts”的适当环境;而翻译环境 C(运动字典)则被记录为“www.yahoo.com/Arts/Recreation/Sports/”的适当环境。数据库可以由系统设计者根据一般用户的使用倾向预先准备好。进一步研究用户的访问习惯可以扩大该数据库。

当用户将某个 URL 字符串输入到某个全球网浏览器的屏幕上



时,借助互联网,全球网浏览器(或该浏览器的代理)可从由 URL 字符串指定的全球网页中下载文件。根据本发明的计算机处理解释该输入的 URL 字符串。在这里,对 URL 字符串的解释引用了查找某个 URL 字符串的一个处理过程,该字符串最接近(即最类似)该输入的 URL 字符串。更具体地说,一个 URL 字符串被看作部分 URL 字符串的一种组合,而具有最接近输入的 URL 字符串的部分 URL 字符串的组合形式的 URL 字符串在数据库中被查找。例如,当输入的 URL 为“www.ibm.com/”(已在数据库中)时,对应的翻译环境 A(互联网字典)被自动地选择。如果输入的 URL 串为“www.yahoo.com/Arts/Architecture/”(该字符串没有记录在数据库中),翻译环境 B(艺术字典),对应“www.yahoo.com/Arts/”,其部分 URL 字符串最接近输入的 URL 串,因此被自动选择。在根据本发明的计算机处理过程中,URL 字符串可以在全球网页已经被下载之后被解释,也可以和下载过程并行处理。根据本发明的计算机处理过程可以包括全球网浏览器下载全球网页的代理功能。

翻译机使用所选择的翻译环境将下载文件中的外语文本翻译成本地语的文本。例如,经翻译得到的本地语言文本被传送给全球网浏览器并被显示在浏览器屏幕上。这时,对原始的外语文本的原始文本窗口和对本地语言文本的翻译窗口可同时被显示在同一屏幕上(原始文本及其对应的翻译文本都可以被显示)。

当用户判定显示在浏览屏上的翻译结果不正确时,可以选择另一个翻译环境并再次执行自动翻译。当对应没有被登记的 URL 字符串的适当翻译环境被找到时,该环境可以被加到该数据库中。

即,根据本发明提供的一种自动翻译方法和机器,在执行对全球网页的自动翻译之前,根据由某个全球网服务器经互联网提供的某个全球网页地址的 URL 字符串,自动地选择某个翻译环境。由于选择了合适的翻译环境,因此提高了翻译的精度。由于翻译环境的选择并将该环境加入到数据库仅仅是对某个 URL 字符串实现的,只涉及到相当少量的数据,因此,处理过程简单而且系统的结构也便于实现。



在自动翻译之前实现的一种技术,从多个翻译环境中,选择适合于原文内容的一种翻译环境,这种技术在本发明提出申请之前就已经被推荐(见前面提到的日本未审查专利公布)。大多数常规的技术选择一种翻译环境,而不是引用分析原文得到的结果的 URL 字符串。换句话说,常规的技术,包括分析原文的一个过程,无论在其结构还是其用法上都是复杂的系统,和本发明有根本的区别。这些常规的技术不是为翻译全球网页而建立的,从某个不同的方面来看,本发明和不同于本发明的常规技术可以共存。

在结合附图详细介绍本发明的实施例的过程中,本发明的其他目的、特征和优点将变得更加明显。

现在将结合附图介绍本发明的最佳实施例。

图 1 给出了根据本发明的计算机系统 10 的硬件结构简图。

图 2 给出根据本发明的自动翻译系统的系统配置的简图。

图 3 说明翻译环境选择器 130 的结构。

图 4 给出显示在全球网浏览器屏上的 URL 数据库 133。

图 5 给出显示在全球网浏览器屏上的 URL 数据库 133(图 4 的继续)。

图 6 是当选择某个翻译环境时,由 URL 比较段 134 执行的一个程序的流程图。

图 7 给出了本发明的自动翻译系统 100 所附加到的全球网浏览器 150 的示例屏幕。

图 8 表示执行加一个登记项到 URL 数据库 133 的程序的流程图。

A. 硬件环境

图 1 是个示意图,说明适合于实施本发明的计算机系统 10 的硬件结构。现在将介绍各个部分。CPU 11,主控制器,在操作系统(OS)的控制下执行各种程序。例如,CPU 11 可以是 Intel 公司的“Pentium/1xx MHz”。操作系统可以是 IBM 公司的“OS/2”(“OS/2 是 IBM 公司的商标)或者是微软公司的“Windows 95”。

CPU 11 通过三条总线和各种硬件模块互相连接:直接连接到



CPU 11 外部引线的处理器总线 12, 用作局部总线的 PCI(外围部件互连)总线 16 和用作输入/输出总线的 ISA(工业标准结构)总线 18。

处理器总线 12 借助桥电路(主机-PCI 桥)13 和 PCI 总线 16 通信联系。在本实施例中的桥电路 13 包括控制访问主存储器 14 的存储控制器, 以及用于吸收总线 12 和 16 的速度之间的差别的数据缓冲区。

主存储器 14 被用作装入程序的一个区域, 也用作当前执行程序的工作区, 并且一般是由多个 DRAM(动态 RAM)芯片组成的。L2-高速缓存 15, 一种高速存储器, 用于临时存放经常被访问的代码/数据并用来吸收 CPU11 访问主存储器 14 所需的时间, 通常由一个 SRAM(静态 RAM)芯片组成。ROM17 为非易失存储器, 其中, 永久地存放着检测程序(POST), 当系统启动时执行该程序, 还存放着代码组, 用于运作系统 10 的各个硬件部分。

PCI 总线 16 是使系统能进行高速数据传输的总线(宽度为 32 位, 操作频率为 33MHz, 最大数据传输速度为 132KBps)。外围设备, 例如高速驱动的视频控制器 20 和卡总线控制器(图中没有表示)都被连接到 PCI 总线 16 上。

视频控制器 20 是用于实际处理由 CPU11 发出的图像处理命令的专用控制器。视频控制器 20 临时将被处理的图像数据存在屏幕缓冲区(VRAM)21 中, 并且从 VRAM21 中读出数据并输出到显示器 22 上。

PCI 总线 16 和 ISA 总线 18 经桥电路(PCI-ISA 桥)19 互相连接。在该实施例中的桥电路 19 包括 DMA 控制器、可编程中断控制器(PIC)和可编程区间定时器(PIT)。在该实施例中的桥电路还有一个 IDE(集成驱动电路)连接器, 用于连接外部的 IDE 存储设备。IDE 硬盘驱动器(HDD)23 和 IDE CD-ROM 驱动器 24 被连接到 IDE 连接器。

上述的桥电路 13 和 19 符合由 PCI 提供的规格说明, 并且一般采用单个芯片组的方式提供。这种芯片组的例子是“Triton”, 由 Intel 公司销售。



ISA 总线 18(宽度为 16 位,最大数据传输速率为 4 MBps)的数据传输速率比 PCI 总线 16 低,用于连接外围设备,例如键盘/鼠标控制器(KMC)25, I/O 控制器 28 和音频控制器 34,这些设备的驱动速度较低。

键盘/鼠标控制器 25 是一个专用控制器,取出在键盘 26 上输入的电脑数据、代码以及经鼠标 27 输入的座标值。

I/O 控制器 28 是一个专用的控制器,用于驱动软盘驱动器(FDD)29 并经并行口 30 或串行口 32 控制并行或串行数据的交换。通常,打印机 31 连接到并行口 30,而调制解调器 33 则连接到串行口 32。

调制解调器 33 借助模拟公用电话线 50 交换数据。调制解调器 33 将计算机系统 10 中的数字信号转换为模拟信号,并经线路 50 传输模拟信号,并且将由线路 50 接收到的模拟信号转换为数字信号并将其提供给计算机系统 10。在该实施例中的计算机系统 10 可通过拨号连接经线路 50 和服务服务器机器 60 通信(更具体地说,是和服务器机器 60 的调制解调器(没有表示)通信)。“服务器机器”是提供互联网连接服务的“服务提供者”所拥有的服务器机器中的一台(例如最接近操作该系统 10 的用户的一台服务器)。服务器机器 60 一般在互联网 70 上被提供,并执行已经对服务器机器 60 拨号的计算机系统 10 对互联网 70 的 IP 联接。例如,服务提供者有 IBM 日本公司的“IBM 互联网连接服务”和 Bekkoame 互联网的“BEKKOAME/INTERNET”。

音频控制器 34 也是一个专用的控制器,用于输入/输出音频信号,音频控制器 34 从麦克风 35 取出计算机数据音频信号,或者执行音频数据的 DA 转换并经喇叭 36 输出其结果。

常用的所谓个人计算机足以用作图 1 所示的计算机系统 10。除图 1 所示的电路外还需很多其他电路来构成计算机系统 10。因为这些电路对本领域技术人员是熟知的,故这里略去对其的说明。

B. 自动翻译系统

图 2 是一个简图说明实施本发明的自动翻译系统 100 的系统配



置。自动翻译系统 100 实际上是作为由图 1 所示的计算机系统 10 运行的应用程序中的一个实现的。

自动翻译系统 100 包括翻译代理 110、翻译机 120 和翻译环境选择器 130。这些模块和全球网浏览器 150 交互作用,更好地执行自动翻译处理过程。

全球网浏览器 150 是一个计算机程序,其作用是互联网的导游。全球网浏览器 150 为计算机系统 10 的用户提供一个用户界面,用于输入 URL(均匀资源定位),即所需全球网页的地址。另外,全球网浏览器 150 解释包含在该全球网页中的 HTML 文件的内容,并且在显示器 22 上显示解释的结果。本实施例中的全球网浏览器 150 具有代理功能。更具体地说,全球网浏览器 150 要求翻译代理 110 从全球网页中下载文件。全球网浏览器 150 还接收来自翻译代理 110 的文本文件的翻译结果,并将其和图像文件(例如 GIF 文件)显示在显示器 22 上。全球网浏览器 150 支持某些显示方式,例如“只显示原本”、“只显示翻译文本”和“原始和翻译文本”,并且这些方式可以在浏览屏上切换。浏览器屏可以是多窗口的一个窗口,后面将介绍。

一旦接到来自全球网浏览器 150 的请求,翻译代理 110 借助互联网 70,从由 URL 指定的某个全球网页中下载一个文件,并将被下载文件中的文本传送给翻译机 120。翻译代理 110 也接收来自翻译机 120 的翻译文本并将其送到全球网浏览器 150 上。“代理”就是某个主机的代理人,该主机不能执行直接的外部访问,“代理”也是中继功能的一般术语,例如接收互联网的通信数据。

翻译机 120 是执行翻译处理过程的主体,即提供原始文本(例如英文这样的外文写的文本)的翻译文本(例如采用日文这样的本地语言写的文本)。利用某个翻译环境,翻译机 120 翻译已经从翻译代理 110 中接收到的原始文本,并将翻译的文本送给翻译代理 110。本实施例中的翻译机具有多个翻译环境,例如翻译环境 A、翻译环境 B 和翻译环境 C。“翻译环境”是用于翻译过程的字典数据库和文法算法(例如设置子句、设置辅助动词的意义,指定句子的风格)。正如通常所知道的,在原文的类型、领域和应用与翻译环境之间存在着相互的



关系。在本实施例中,为每个类型提供特定的字典数据库,例如互联网字典、商业字典、政治字典、娱乐字典、艺术字典或运动字典。

翻译环境选择器 130 根据需要改变由翻译机 120 使用的翻译环境。在本实施例中,翻译环境选择器 130 的设计基于这样的事实:URL(某个全球网页的地址)是由若干字符串段(即“部分 URL 字符串”)组成的,这些字符串段分别描述协议名、服务器名和文件名。翻译环境选择器 130 根据全球网浏览器 150 输入的 URL 字符串,自动选择某个适当的翻译环境。

图 3 是一个简图,说明翻译环境选择器 130 的结构。如图 3 所示,翻译环境选择器 130 被划分为 URL 输入段 131、翻译环境输出段 132、URL 数据库 133、URL 比较段 134 和基本操作段 135。

URL 数据库 133 是一个表格,在该表中给出了被相当频繁访问的全球网页的 URL 字符串,以及在每个 URL 字符串及其适当的翻译环境之间的对应关系。

图 4 和图 5 给出了显示在全球网浏览器屏上的 URL 数据库 133(图 5 是图 4 中屏幕显示的继续)。如图 4 和图 5 所示,被经常访问的全球网服务器/全球网页的 URL 字符串被登记在数据库 133 中,并列出了与其适当的翻译环境的相互关系。例如,“www.ibm.com/”和“www.yahoo.com/”和它们的适当的翻译环境“互联网,一般的”结合在一起被登记,而“cnn.com/”则与其适当的环境“政治”结合在一起被登记。在本实施例的 URL 数据库 133 中,各字符串不规则地被描述(即,只按照它们被登记的次序描述,而不是按照其他的任何次序,例如按字母顺序、部分 URL 字符串排列或串的计数)。

涉及 URL 字符串和其合适的翻译环境之间对应关系的一个特点是:当 URL 字符串描述相同的全球网服务器名而被指定的文件不同时,在许多情况下,适合的翻译环境也是不同的。如图 4 和图 5 所示,例如,“互联网,一般的”被登记作为“www.yahoo.com/”的一个合适环境。然而,“政治”、“运动”和“艺术”被登记为同一全球网服务器中的分别为“www.yahoo.com/ Government/”、“www.yahoo.com/Recreation/Sports/,”和“www.yahoo.com/Arts/”的文件的合



适的环境。另外,根据最初由本发明人设计的经验规则,发现即使在同一全球网服务器中的文件中,具有最相互接近或最类似的“部分 URL 字符串”(在构成一个 URL 的字符串中由“/”分开的字符段)的组合形式的文件倾向于利用相同的合适的翻译环境。例如,由 URL 字符串“www.yahoo.com/Arts/Architecture”指定的文件具有适当的翻译环境“arts”,这和字符串“www.yahoo.com/Arts/”的相同,其部分 URL 字符串比“www.yahoo.com/”的更接近。

自动翻译系统 100 的设计者可以预先在 URL 数据库 133 中登记 URL 字符串。此外,系统 100 的用户也可以借助基本操作段 135 来登记 URL 字符串,这将在下面介绍。

通过翻译代理 110,URL 输入段 131 接收在全球网浏览器屏幕上输入的某个 URL 字符串,并将其传送给 URL 比较段 134。URL 比较段 134 执行一个比较过程,以便在 URL 数据库 133 中输入 URL 字符串,并找出某个适当的翻译环境。URL 比较段通过翻译环境输出段 132 向翻译机 120 报告比较结果。

C. 翻译环境的选择

URL 比较段 134 在 URL 数据库 133 中查找某个 URL 字符串,该串最接近输入的 URL 字符串,并将与得到的 URL 字符串对应的某个翻译环境输出到翻译机 120。图 6 是当选择翻译环境时,由 URL 比较段 134 执行的程序的流程图。在该程序中,S 表示由全球网浏览器输入的 URL 字符串,i 表示在 URL 数据库 133 上的一个索引,而 m 则表示在查找过程中得到的最接近 S 的 URL 字符串的索引。例如,S(i)表示被登记为 URL 数据库 133 中的第 i 个记录的 URL 字符串。当由数据库查到得到的最后结果是 $m = M$ 时,意味着合适的翻译环境就是对应 URL 数据库 133 的第 M 个 URL 字符串 S(M)的那个翻译环境。E(m)表示对应被登记在 URL 数据库 133 的第 m 个记录中的 URL 字符串的翻译环境。

在该程序中,首先,为每个索引置初值(步 S10)。索引 i 的初值为 1,表示 URL 数据库 133 的第一个记录。索引 m 的初值为 0,表示类似的 URL 字符串还没有找到。



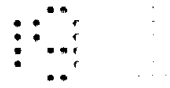
在步 S20 中, 执行检查以确定由当前索引 i 保持的值是否已经超出了 URL 数据库 133 中最后记录的计数。

当判定步骤 S20 得到的结果为否定时, 即, 当某个没有被比较的记录仍留在 URL 数据库 133 中时, 在步 S30 中执行检查以判定当前输入的 URL 字符串 S , 作为一个部分 URL 字符串, 是否包括第 i 个记录的 URL 字符串 $S(i)$ 。

当判定块 S30 得到的结果为否定时, 即, 当 URL 字符串 $S(i)$ 并不类似当前输入的 URL 字符串 S 时, 程序控制跳到步 S70, 在步 S70 中, i 被增 1, 以便比较 URL 数据库 133 中的下一个记录和当前的 URL 字符串 S 。然后将程序控制返回到步 S20。

当判定模块 S30 得到的结果为肯定的时, 即, 当 URL 字符串 $S(i)$ 接近当前输入的 URL 字符串 S 时, 程序控制跳到步 S40。在步 S40 中, 执行检查以判定是否 $m=0$, 即, 是否类似的字符串还没有被找到。当 $m=0$ 时, i 替换 m , 以便临时存放用于比较的 URL 字符串。

在判定块 S40 中, 如果 $m \neq 0$, 即, 如果某个类似的 URL 字符串已经被临时存放起来, 程序控制移到步 S50。在步 S50 中, 将当前正用于比较的 URL 字符串 $S(i)$ 和已经被存放起来的 URL 字符串 $S(m)$ 进行比较, 以便确定该 URL 字符串是否更接近当前输入的 URL 字符串 S 。实现这一判定是通过比较 URL 字符串 $S(i)$ 和 $S(m)$ 以判定哪一个具有更大的部分 URL 字符串的序号。对于“www.yahoo.com/”和“www.yahoo.com/Arts/”来说, 由于前者包含一个部分 URL 字符串, 而后者包含两个, 因此后者是更好的。所谓“更大的部分 URL 字符串的序号”是指: 即便更多的部分 URL 字符串互相对应, 除非部分 URL 字符串的排列不匹配, 作为 URL 比较的结果的相似性就低。假定当前输入的 URL 字符串为“www.yahoo.com/Recreation/Sports/”, 并且有两个 URL 字符串“www.yahoo.com/Sports/”和“www.yahoo.com/Recreation/”被登记在 URL 数据库 133 中。虽然后者有两个顺序部分 URL 字符串匹配输入的 URL 字符串, 而前者只有一个部分 URL 字符串匹配输入的



URL 字符串(虽然“Sports”被包含在输入的 URL 字符串中,但不能计数,因为其排列不能互相对应),因此,后者的 URL 字符串被优选。

在判定块 S50 中,当确认新比较的 URL 字符串 $S(i)$ 较好时,在步 S60 中, i 替换 m ,以便改变最类似的 URL 字符串。接着,程序控制进入步 S70。

在判定步 S50 中,如果先前被存放的 URL 字符串 $S(m)$ 较好,则程序控制跳到步 S60,维持 m 的当前值,并进入步 S70。

在步 S70 中, i 被增 1 以便比较 URL 数据库 133 中的下一个记录,并且接着将程序控制返回步 S20。在步 S20 中,如果在 URL 数据库 133 中还有没有被比较的记录,则重复步 S30 到 S70 的处理过程。在步 S20 中,当确认当前索引 i 保持的值已经超出 URL 数据库 133 中最后记录的计数时,程序控制分支到“是”并转到步 S80。

在步 S80 中,执行检查以判定是否 $m=0$ 。当 $m=0$ 时,意味着在 S20 到 S70 的循环中,在 URL 数据库 133 中没有找到接近当前输入的 URL 字符串 S 的一个 URL 字符串。在这种情况下,将缺省的翻译环境输出到翻译机 120(步 S90),并终止选择翻译环境的程序。

如果 $m \neq 0$,作为在循环 S20 到 S70 中被认为最类似的 URL 字符串对应的翻译环境 $E(m)$ 被输出到翻译机 120(步 S100)。接着终止选择翻译环境的程序。翻译机 120 使用所选择的翻译环境 $E(m)$ 翻译下载的文件。

D. 更新 URL 数据库

如上所述,URL 数据库 133 的内容通过基本操作段 135 被另外登记。在本部分中,增加某个新登记项到 URL 数据库的过程将被简单介绍。

图 8 是增加一个新的登记项到 URL 数据库 133 的程序的流程图。首先,将翻译的文本显示在全球网浏览器屏幕上,询问用户对该翻译结果是否满意,即,是否已经选择了一种合适的翻译环境(步 S200)。

当判定块 200 得到的结果是否定的时,不需要增加一个登记项,



并且跳过所有下面的步骤,该程序被终止执行。

如果判定块 S200 得到的结果是肯定的,询问用户是否将新输入的 URL 字符串 S 及其合适的翻译环境之间的关系加到 URL 数据库 133(步 S210)。用户可根据将来对该新输入的 URL 字符串 S 的引用频率来决定是否执行登记工作。

如果在判定块 S210 中得到的结果是肯定的,将这种关系登记在 URL 数据库 133(步 S220)。接着终止程序执行。如果判定块 S210 的结果是否定的,则将程序控制跳到步 S220 并终止该程序。

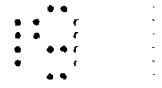
E. 浏览器屏幕

图 7 是本发明的自动翻译系统 100 所附加到的全球网浏览器 150 的屏幕。图 7 的屏幕为“原始和翻译文本显示”方式,在这个方式下,被翻译的文本直接显示在原文之下。这种方式可以被切换到另一种方式,在全球网浏览器屏上“只显示原文”或“只显示翻译文本”。全球网浏览器屏可以是显示器 22 上多个窗口中的一个。

从顶部读下来,在全球网浏览器屏上有“标题条”、“菜单条”和“工具条”。众所周知,标题条是显示在该屏幕上被激活的某个应用软件名称的区域,而菜单条是显示可以被选择的一系列命令的区域。同样众所周知的是,工具条是菜单条的一种类型,其中被排列的按钮用于快速激活常用的功能。这些按钮也被称为“灵巧图标”。

紧接在工具条下面的是 URL 输入部。用户可以在 URL 输入段中键入某个 URL 字符串以指定将被访问的所需全球网页。在图 7 中,URL 名“www.yahoo.com/”被输入。

输入的 URL 字符串被送到翻译代理 110。翻译代理 110 代替全球网浏览器 150 从被指定的全球网页中下载文件,并将该文件中的原始文本临时送到翻译机 120。借助翻译代理 110 将输入的 URL 字符串送到翻译环境选择器 130 中。翻译环境选择器 130 解释所接收到的 URL 字符串并为其选择一个合适的翻译环境,该翻译环境被报告给翻译机 120。翻译机 120 使用所选的翻译环境翻译原始文本并将原文和翻译的文本通过翻译代理 110 送到全球网浏览器 150。这一处理系列是在桌面机的背景中实现的。



在 URL 输入段下面的一个矩形区域, 占据了浏览屏的大部分, 是全球网浏览器屏幕的主体, 其中所显示的是被下载全球网页的内容。在图 7 中, 由 URL 名指定的“www.yahoo.com/”页被显示在该屏上。

在全球网浏览器屏的主体的最上方给出了一个指示器, 用于显示当前的显示方式和所选的翻译环境。在图 7 中, “原文和翻译文本”方式被高亮度显示为当前的显示方式, 因此, 被翻译的文本直接显示在屏幕上的原文之下。如图 4 所示, 由于“互联网, 一般的”被登记为 URL 数据库中的“www.yahoo.com/”的适当环境, 有关的消息被显示在指示器中。用户可以动态地改变显示方式和所用的翻译环境, 只需用鼠标点该指示器。

F. 附录

本发明在引用具体的实施例时已经被详细介绍。然而, 对于具有该技术的一般知识的人来说, 显然能在本发明的范围内对该实施例进行各种修改和修正。例如, 自动翻译系统并不局限于使用全球网浏览器的代理功能的系统, 而可以是借助某个剪辑板在全球网屏上接收原始文本并翻译的系统。即, 虽然本发明的提出是利用一个例子, 但不应该局限于该例子。为了充分理解本发明的主题, 应该参考权利要求书。

如上所详述的, 根据本发明, 提供一种优良的自动翻译方法和机器, 利用装有多个翻译环境的翻译系统, 将外文的文本(例如英文)翻译成本地语言文本(例如日文, 朝鲜文或中文文本)。

此外, 根据本发明, 提供的一种优良的自动翻译方法和机器, 在将外语文本翻译成本地语言文本之前, 为将被翻译的外语文本自动选择一种适当的翻译环境。

此外, 根据本发明, 提供一种优良的自动翻译方法和机器, 使用装有多个翻译环境的翻译系统, 将某个全球网服务器经互联网提供的某个全球网页中的外语文本正确地翻译成本地语言的文本。



URL 字符串	翻译环境
www.ibm.com/	翻译环境 (互联网字典) A
www.yahoo.com/	翻译环境 (互联网字典) A
www.yahoo.com/Arts/	翻译环境 (艺术字典) B
www.yahoo.com/Arts/Recreation/Sports/	翻译环境 (运动字典) C

表 1

说明书附图

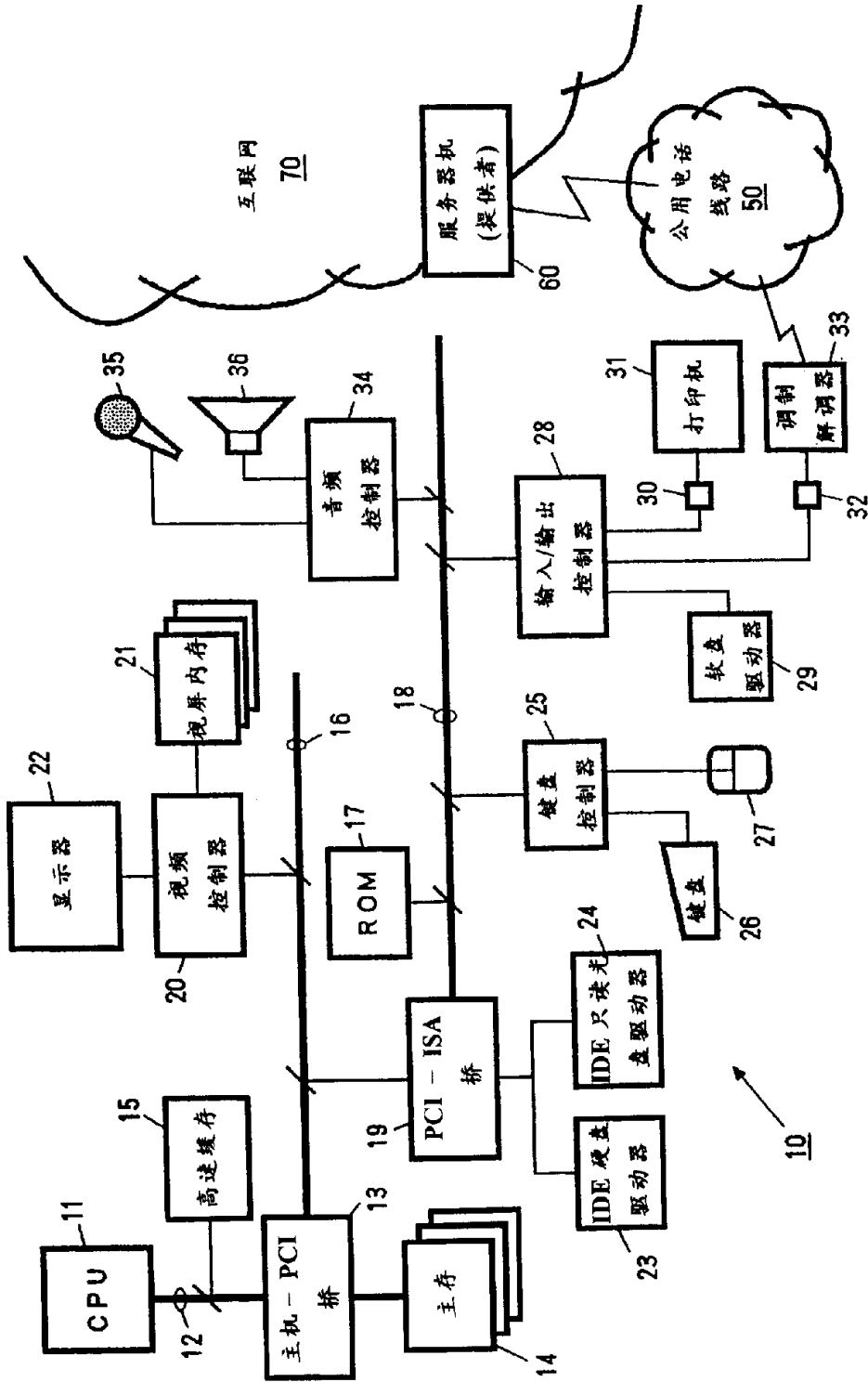


图 1

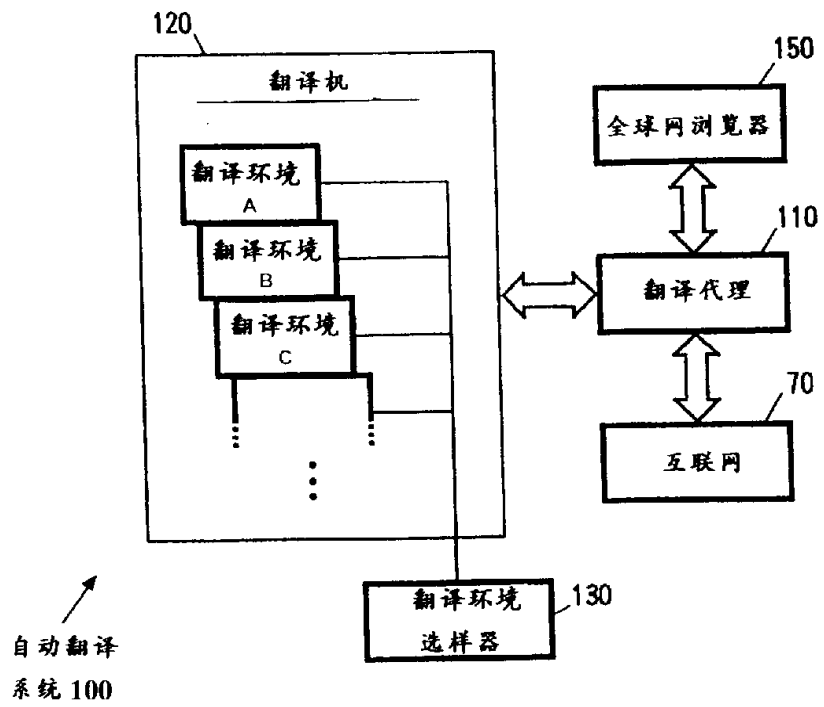


图 2

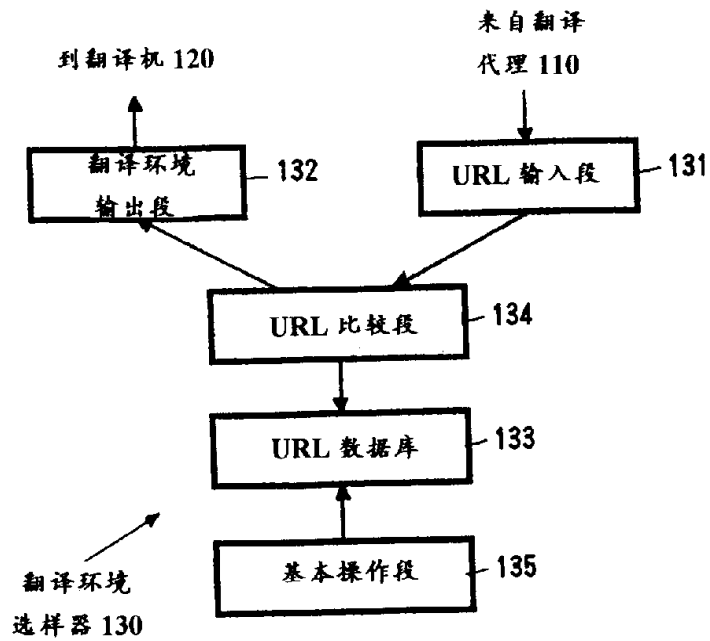


图 3

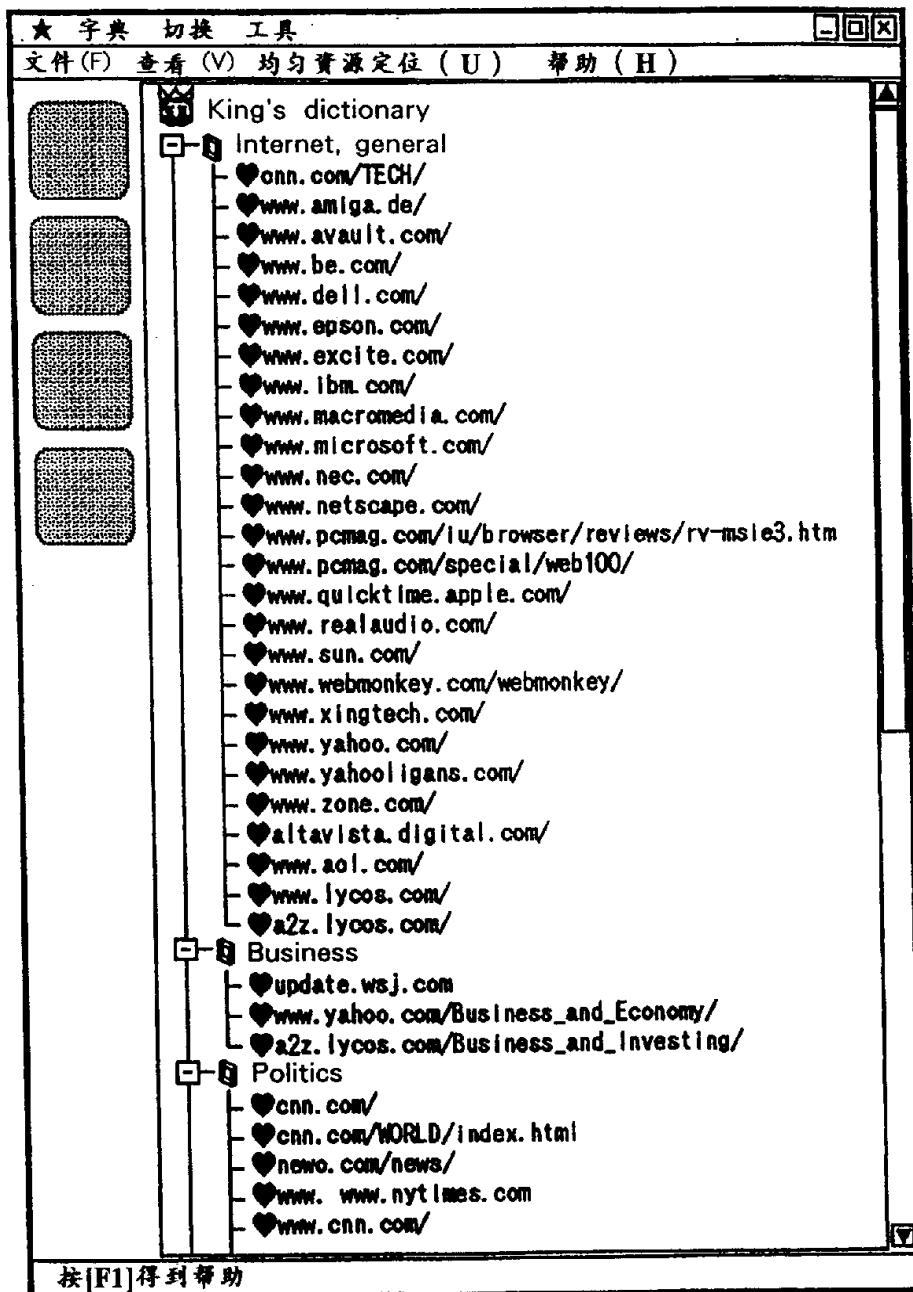


图 4

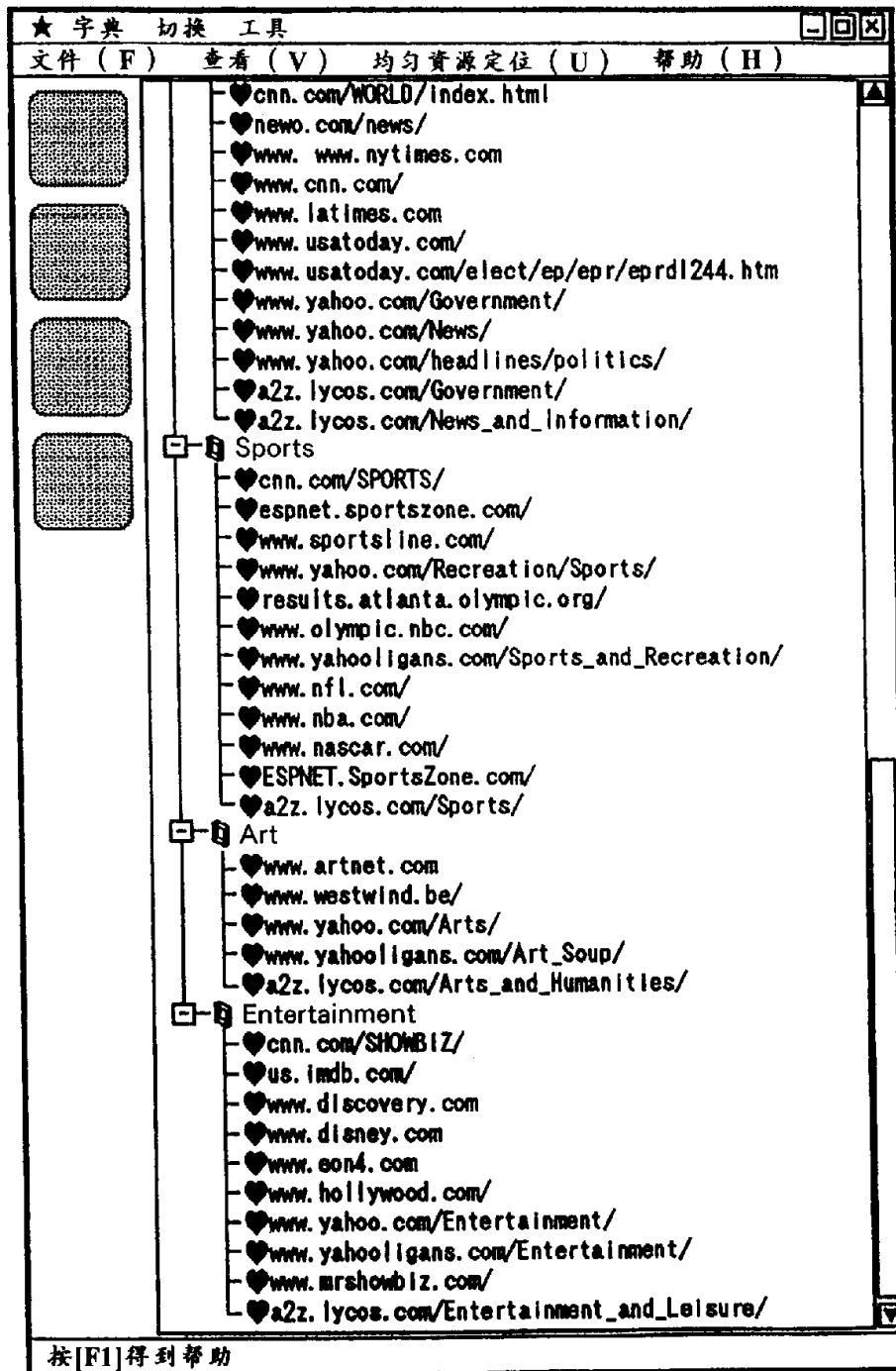


图 5

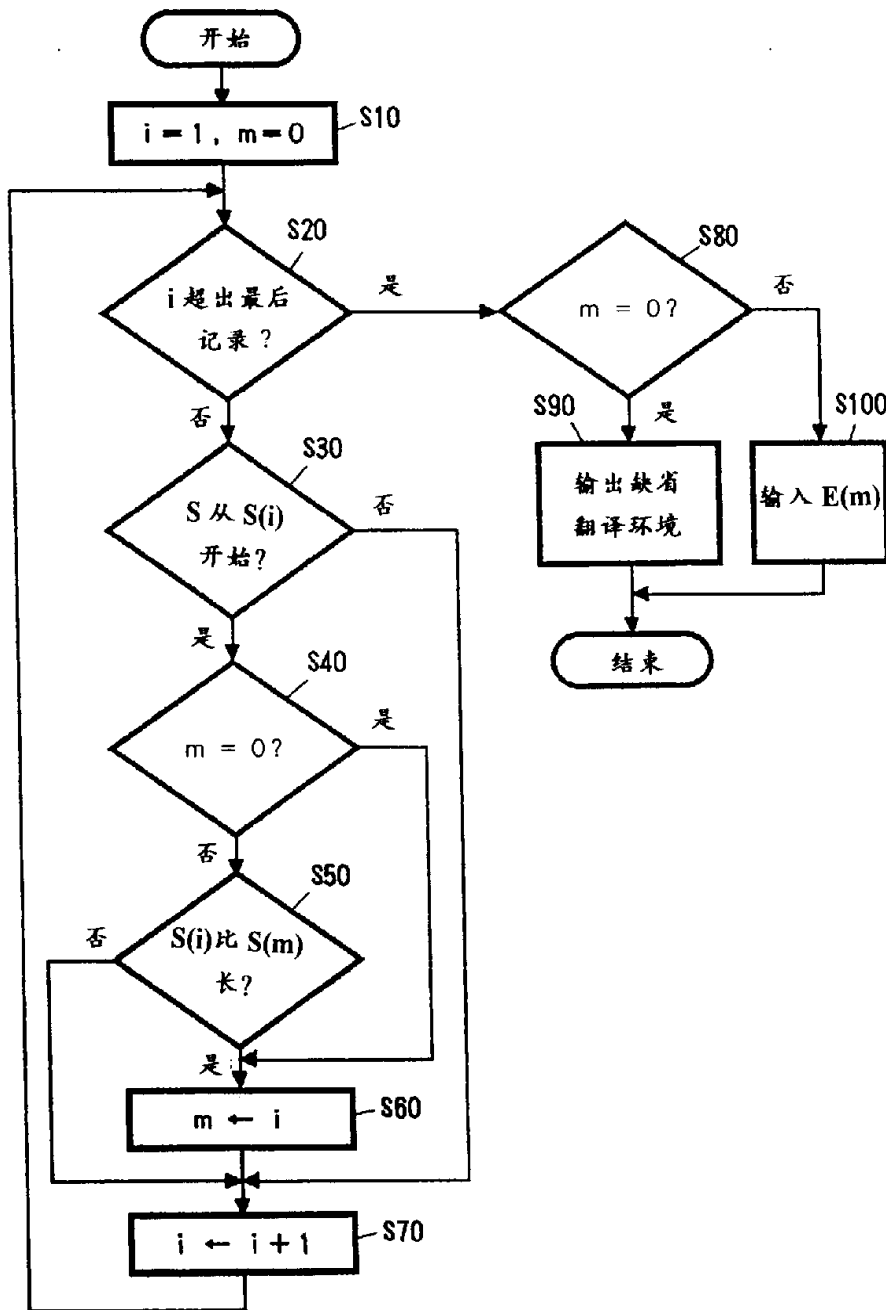
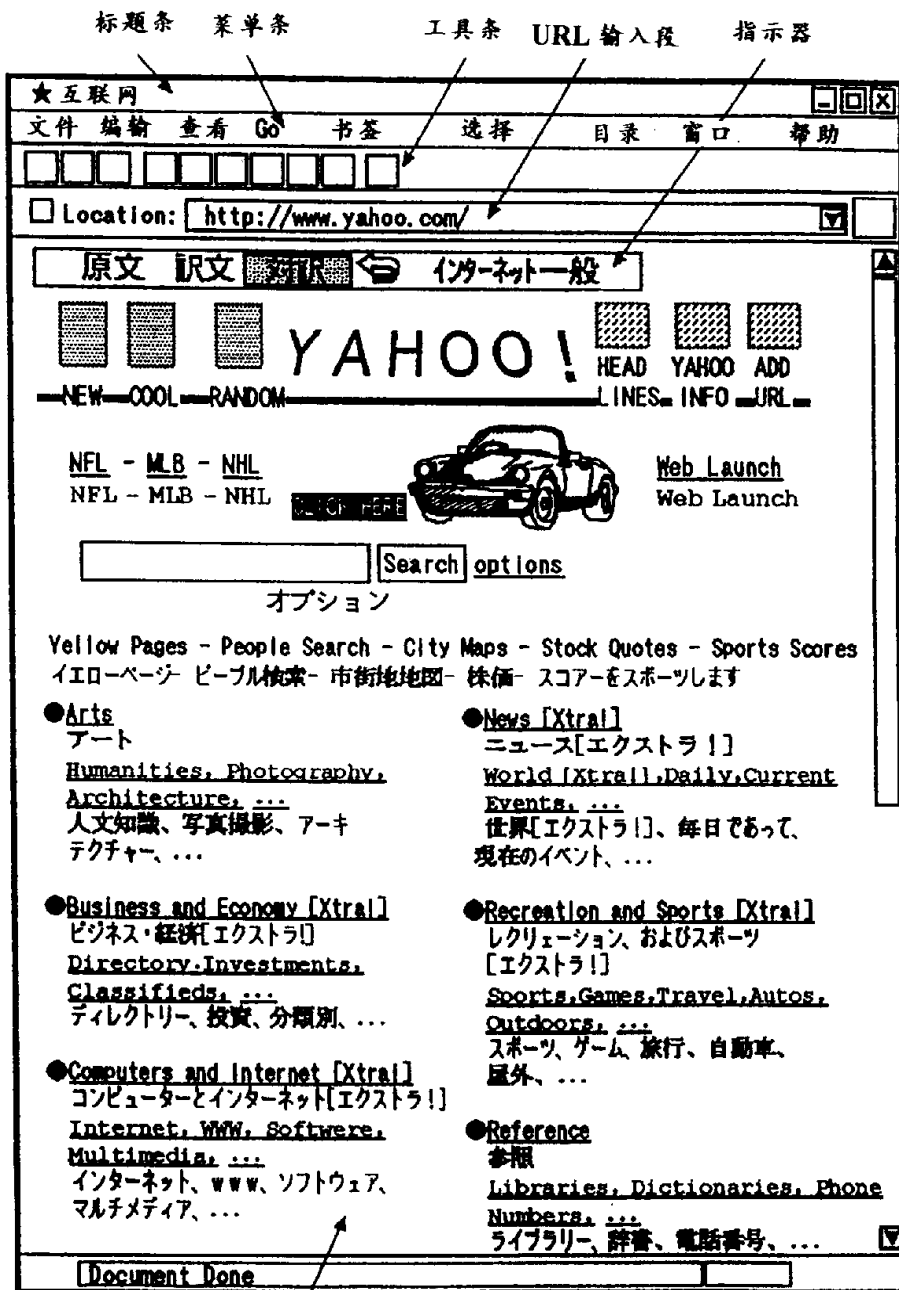


图 6



标题条 菜单条 工具条 URL 输入段 指示器

浏览器屏幕

图 7

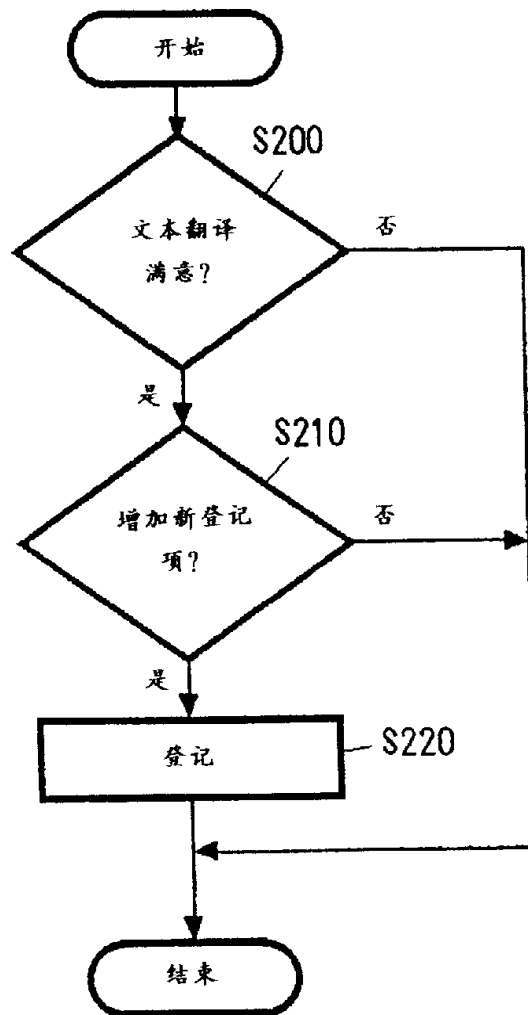


图 8