



[12] 发明专利申请公开说明书

[21] 申请号 02805227.7

[43] 公开日 2004 年 4 月 28 日

[11] 公开号 CN 1493043A

[22] 申请日 2002.2.21 [21] 申请号 02805227.7

[30] 优先权

[32] 2001.2.22 [33] AU [31] PR3295

[86] 国际申请 PCT/AU2002/000184 2002.2.21

[87] 国际公布 WO02/067139 英 2002.8.29

[85] 进入国家阶段日期 2003.8.20

[71] 申请人 世界链接公司

地址 美国内华达州

[72] 发明人 P·斯坎伦

[74] 专利代理机构 北京纪凯知识产权代理有限公司

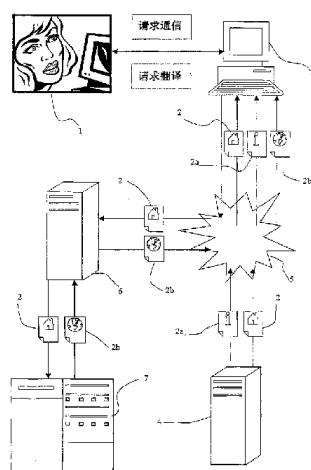
代理人 沙 捷

权利要求书 3 页 说明书 16 页 附图 3 页

[54] 发明名称 翻译信息段

[57] 摘要

一种利用翻译信息段来提供网络环境中通信信息的无缝翻译的方法和设备。该通信信息包含一可读的翻译信息段，该翻译信息段包括全局翻译参数以获得从第一语言到目标语言的通信信息的翻译。翻译信息受到分析以获得翻译该通信信息时所使用的翻译参数。翻译信息最好包括以下至少其中之一：指向通信信息翻译的一指示字；或是指向用于对通信信息进行翻译的翻译引擎的一指示字；或是指向可熟练翻译电子通信信息的翻译者的一指示字。



1. 一种与电子通信信息相关的翻译信息段：

所述翻译信息段包括全局变量，用以将所述电子通信信息或其一部分或多个部分从源语言翻译成一种或多种目标语言；

- 5 所述翻译段受到读取该电子通信信息的应用程序的识别和作用，
以析取翻译参数，从而获得从所述源语言到所述一种或多种目标语言
的该电子通信信息的翻译。

2. 根据权利要求 1 的翻译信息段，其中该翻译信息段被嵌入在该
电子通信信息中或附加在该电子通信信息上。

- 10 3. 根据权利要求 1 的翻译信息段，其中该翻译信息段被存储在可
存取的数据库中，并且有一个指示字或多个指示字被嵌入在或附加在
可翻译的电子通信信息中。

15 4. 根据权利要求 1 至 3 中任一项的翻译信息段，其中该全局参数
选自以下参数：源语言；编码；时态；可用的翻译；翻译引擎；词典；
词汇；上下文；翻译服务；对标记符进行处理的规则；对诸如图片、
图形、声音、动画、视频、软件、编程例程这样的电子通信信息内组
分进行处理的规则；执行翻译的规则；现有翻译的位置；诸如图片、
图形、声音、动画、视频、软件、编程例程、个人翻译者、以及翻译
存储器这样的所述电子通信信息的现有本地组分的位置。

20 5. 根据权利要求 1 至 4 中任一项的翻译信息段，其中该 TIS 的两
个或多个参数相互配合以翻译所述电子通信信息的一部分或多部分。

6. 根据权利要求 1 至 5 中任一项的翻译信息段，其中作用于该翻
译信息段的应用程序包括一用于网页的万维网浏览器、或一用于电
子邮件的电子邮件程序、或一用于文本文档的字处理器。

25 7. 根据权利要求 1 至 5 中任一项的翻译信息段，其中作用于该翻
译信息段的应用程序是一个检测并作用于该翻译信息段的特殊目的应
用程序。

8. 根据权利要求 1 至 7 中任一项的翻译信息段，其中存在两个或更多个与所述电子通信信息相关的翻译信息段。

9. 根据权利要求 8 的翻译信息段，其中每个翻译信息段均包括用于对相关于该翻译信息段的电子通信信息部分进行翻译的参数。

5 10. 一种与电子通信信息相关的翻译信息段，所述翻译信息段受到读取该电子通信信息的应用程序的识别和作用，并且所述翻译信息段包括以下至少其中之一：指向该电子通信信息的翻译的一指示字；指向现有翻译位置的一指示字；指向诸如图片、图形、声音、动画、视频、软件、编程例程这样的所述电子通信信息的现有本地组分的位置的一指示字；指向用于执行翻译的规则的一指示字；指向用于对诸如图片、图形、声音、动画、视频、软件、编程例程这样的电子通信信息内的组分进行处理的规则的一指示字；指向用于对该电子通信信息进行翻译的翻译引擎的一指示字；指向词典、词汇、或术语数据库的一指示字；或者指向可熟练翻译该电子通信信息的翻译者的一指示字。

11. 根据权利要求 10 的翻译信息段，其中指向该电子通信信息的翻译的指示字是通用资源定位器和指向不同语言翻译的指示字列表。

12. 根据权利要求 10 的翻译信息段，其中该翻译信息段包括一个翻译参数的列表或者包括一个指向包含有翻译参数列表的文件的指示字。

13. 根据权利要求 12 的翻译信息段，其中该翻译参数可由翻译引擎或人工翻译者读取以提高翻译质量。

14. 一种为外语通信信息的接收者提供被翻译的通信信息的方法，该方法包括步骤：

25 使一翻译信息段与该外语通信信息相关联；
将该外语通信信息与翻译信息段传送到接收方；
对该外语通信信息进行分析以识别并分析该翻译信息段；以及
根据该翻译信息段中的参数而获得该外语通信信息的翻译。

15. 根据权利要求 14 的为外语通信信息的接收者提供被翻译的信息的方法，其中由一浏览器请求翻译，且从该通信信息中析取该翻译信息段的信息，并将其与翻译请求一起传送到一翻译管理器。

16. 根据权利要求 14 的为外语通信信息的接收者提供被翻译的信息的方法，其中：

当浏览器接收到要显示的通信信息时，它首先检查该翻译信息段以在显示之前确保语言是正确的；并且

如果所述语言是不正确的，则该浏览器向翻译管理器请求翻译。

17. 根据权利要求 14 的为外语通信信息的接收者提供被翻译的信息的方法，其中：

网络服务器获得用户的优选语言并将其与该翻译信息段进行比较，而如果它不匹配，则

该网络服务器请求对该通信信息进行翻译并将来自该翻译信息段的相关细节提供给该翻译管理器。

翻译信息段

本发明一般地涉及翻译领域，更具体地说，涉及有便于通信信息
5 的无缝翻译的翻译信息段。

本发明还涉及一种利用翻译信息段来对网络环境中的通信提供无
缝翻译的方法和装置。

发明背景

10 将通信信息从一种语言到另一种语言的机器翻译正在打破个人之
间与企业之间的通信障碍。在过去的二十年中，已稳定改善了机器翻
译的质量。已经研发出各种按照短句而非逐词进行翻译的技术。其他
技术是利用词典或翻译存储器来翻译整个句子。其结果是，改善了被
译信息的语法并且由此提高了其可读性。对于通用语言以及特定目的
15 而言，一些较好的翻译程序已接近人工翻译的质量。

尽管已明显了提高了机器翻译软件的技能，但是几乎没有提高其
适用性。为了翻译文档、电子邮件、或其他的通信信息，通常必须要
访问翻译站点并且运行一翻译程序。用户必须输入诸如源语言和目标
语言、优选词典、以及特殊词等等这样的程序参数。

20 在我们共同待审的申请号为 09/6766690 的美国专利申请中，我们
描述了一次点击翻译系统，该系统可避免大部分过去为获得通信信息
的翻译所必须的用户输入。一次点击翻译系统包括一个一次点击翻译
组件以及一个翻译管理器，一旦用户点击了一次点击组件，上述两者
就共同提供近似无缝的翻译。

25 尽管对于现有技术而言一次点击系统是一个显著进步，但是它仍
需要通信信息的接收方起一定的作用。为了对通信信息进行翻译以使
其能被普遍地接受，它必须完全无缝。这要求系统自动按照接收者的
优选语言来传送通信信息。系统还必须传送比目前所提供的质量更好的
翻译。

30 一些最新的技术取得了进步，但未能实现该设想。例如，转让给

AT&T 公司的公开号为 6161082 的美国专利描述了一种网络基语言翻译系统，该系统旨在通过利用网络的处理能力而非本地机来执行翻译，从而改善机器翻译。然而，该专利未能清楚阐明如何检测所涉及的语言。它只述及可从两方之间的通信中检测出源语言和目标语言，而没有表明这是怎样完成的。AT&T 方法未能对翻译质量提供任何本质性提高，改善仅仅是由网络环境中可用的增加的处理能力来实现的。
5

申请人为富士通有限公司、公开号为 5548508 的美国专利旨在通过将标记符嵌入在包括有上下文信息的文档中来改善机器翻译的质量。例如，<TITLE>...</TITLE>标记符表示该词是标题并且会被相应显示，<MODIFY>...</MODIFY>标记符被用于定义所译词的正确顺序。上述标记作用于一部分文档而不是全部。为了使其有效，该发明需要一个可支持富士通扩展标识符集合的翻译程序。富士通发明实现了提供具有较高精确度的机器翻译这一目的，但是该发明为此付出了会减慢翻译速度的预处理和后处理的重大代价。利用富士通方案不可能以无缝方式提供机器翻译。
10
15

最近授权的公开号为 6073143、转让给为三洋电子有限公司的美国专利描述了一种通过将翻译命令添加到文档中的每个超级链接来强化 HTML 文档翻译的方法。该发明寻求解决在翻译过程中所出现的丢失超级链接的问题。它仍未改善实际文档的翻译。

公开号为 5848386、转让给理光公司的美国专利描述了一种自动翻译系统，其使用诸如词典和规则数据库之类不同翻译资源来翻译文档的不同部分。标记符被嵌入在文档中以定义待译文档的结构，以便选择翻译处理过程中要使用的词典和/或规则。但该系统要求每个对待译文档生成“文档类型规定”。翻译仅作用于具有预定结构的文档，而不是诸如因特网上的网页之类的文档。
20
25

发明内容

一方面——尽管其未必是唯一的或者实际上最广的方面，但是本发明在于与电子通信信息有关的翻译信息段：

所述翻译信息段包括全局变量，用以将所述电子通信信息或其一部分或多部分从源语言翻译成一种或多种目标语言；并且
30

所述翻译段被读取电子通信信息的应用程序所识别和作用，以析取翻译参数，从而获得从所述源语言到所述一种或多种目标语言的电子通信信息的翻译。

可将翻译信息段嵌入在电子邮件通信信息中或将其附加在电子通信信息上。或者，将翻译信息段存储在可存取的数据库中，并且将一个指示字或多个指示字嵌入在或附加在可译的电子通信信息中。
5

所述全局参数选自以下参数，其包括但并不限于：源语言；编码；时态；可用的翻译；翻译引擎；词典；词汇；上下文；翻译服务；单个翻译者；对诸如 HTML 标记符这样的标记符进行处理的规则；对诸如图片、图形、声音、动画、视频、软件、编程例程这样的电子通信信息内的组分进行处理的规则；执行翻译的规则；现有翻译的位置；诸如图片、图形、声音、动画、视频、软件、编程例程以及翻译存储器这样的所述电子通信信息的现有本地组分的位置。
10

最好是，作用于翻译信息的应用程序包括用于网页的万维网浏览器、或用于电子邮件的电子邮件程序、或用于文本文档的字处理器。
15 或者，特殊目的（purpose specific）应用程序可检测并作用于所述翻译信息段。

可存在两个或更多个与电子通信相关的翻译信息段。每个翻译信息段均包括参数，用以对相关于翻译信息段的电子通信部分进行翻译。
20

在另一方面，本发明在于与电子通信相关的翻译信息段，所述翻译信息段受到读取电子通信的应用程序的识别和作用，并且所述翻译信息段包括以下至少其中之一：指向要翻译的电子通信信息的一指示字；指向现有译文位置的一指示字；指向诸如图片、图形、声音、动画、视频、软件、编程例程这样的所述电子通信信息的现有本地组分位置的一指示字；指向用于执行翻译的规则的一指示字；指向用于对诸如图片、图形、声音、动画、视频、软件、编程例程这样的电子通信信息内的组分进行处理的规则的一指示字；指向用于翻译电子通信信息的翻译引擎的一指示字；指向词典、词汇、或术语数据库的一指示字；或者指向可熟练翻译电子通信的人工翻译者的一指示字。
25

指向电子通信信息翻译的指示字适合是通用资源定位器且最好是
30 指向不同语言翻译的指示字列表。

翻译信息段最好还包括一个翻译参数的列表或者包括一个指向包含有翻译参数列表的一文件的指示字。翻译参数适合由翻译引擎或人工翻译者读取以提高翻译质量。

在又一个方面，本发明在于一种方法，其用于为外语通信信息的
5 接收者提供经过翻译的通信信息，该方法包括以下步骤：

使翻译信息段与外语通信信息相关联；

将外语通信信息与翻译信息段传送到接收者；

对外语通信信息进行分析以识别并分析翻译信息段；以及
根据翻译信息段中的参数而获得外语通信信息的翻译。

10 当由浏览器请求翻译时，可从通信信息中析取翻译信息段的信息
并将其与翻译请求一起传送到一翻译管理器。

或者，当浏览器接收到将显示的通信信息时，它可首先检查翻译
信息段以在对其显示之前确保语言是正确的，不然的话，它可向翻译
管理器请求翻译。

15 另一选择是使网络服务器获得用户的优选语言并将其与翻译信息
段进行比较，若其不匹配，则网络服务器请求对通信信息进行翻译并且
将来自翻译信息段的相关细节提供给翻译管理器。

另外又一选择是使机器翻译引擎直接察看翻译信息段，并且利用
该信息以实现较佳翻译。

20 翻译信息段信息用在客户端或者服务器端是无关紧要的。关键是
翻译信息段内的信息是用来帮助翻译管理器来获得最好的翻译。

在一些情况下，可省略翻译管理器。例如在浏览器察看翻译信息
段何时请求翻译、并察看 URL 何处可得到所希望的翻译时，该浏览器
可以直接由所述 URL 请求翻译。

25

附图简要说明

为了有助于理解本发明，参考随后的附图来对优选实施例进行详细说明。

30 图 1 给出了无缝翻译处理的流程图；
图 2 给出了非无缝翻译处理的流程图；以及

图 3 给出了利用翻译信息段的翻译处理的系统概观。

附图详细说明

参考图 1, 其中给出了将通信信息从源语言翻译成目标语言的无缝方法的流程图。为便于描述, 就网页将源语言单翻译成目标语言来对该方法进行详细地描述。应该理解的是, 对于将所述处理加以扩展, 以将多个通信信息翻译成多种语言而言, 这是很通常的。另外, 根据所述方法, 可对包括有文本文档、电子邮件、SMS 消息、音频文件、以及视频等等的任意电子通信信息进行翻译。

所述方法的要素是通信信息中包括有一翻译信息段 (TIS)。该 TIS 为通信信息的发送者以及接收者提供信息以进行无缝翻译。在最简单的方面, TIS 提供一段信息以帮助获得通信信息的最佳翻译。这一段信息可以是在诸如 URL 这样的应用程序或者翻译存储器中所描述的任意参数。上述应用程序已具有一个可用通信信息的专业翻译版本。上述翻译存储器已具有多个已译的短语和句子。

TIS 不仅仅是用于获得较好的机器翻译, 而且还用于设法获得最佳的翻译。这意味着可借用现有可能的人工翻译, 因此如果专业的人工翻译版本是可用的, 那么可获得该版本。或者如果翻译存储器中的人工翻译段是可用的, 则可获得人工翻译段。

在更复杂的一方面 (随后对其进行详细描述), TIS 包括所有参数的较详细列表以获得将通信信息翻译成目标语言的较好译文。上述所有参数包括诸如音调、主题、优选词典、优选词汇、优选翻译引擎、要排除的单词、要忽略的数据、翻译服务、现有翻译的位置、诸如图片、图形、声音、动画、视频、软件这样的现有本地组分的位置、对标记符 (例如 HTML 标记符) 进行处理的规则、对诸如图形、指南、编程例程、以及付帐方法 (用于商业翻译) 这样的通信信息组分进行处理的规则。应认识到, 诸如词典或词汇这样的其中一些参数对人工翻译者以及机器翻译器都是有用处的。TIS 的优点并不局限于机器翻译。

TIS 还可择包括一个或多个指向一文件的指示字, 该文件包含了上述列表中的某些或所有参数。这一实施例实质等价于使信息嵌入在通信信息内, 但其对翻译管理器维持翻译参数的数据库这样的网络环境

可能是更有效的，其中上述翻译参数是根据翻译时的 TIS 标识符来检索的。

还可为不包括有 TIS 的通信信息提供一种无缝翻译系统，但是该处理已在发明名称为“无缝翻译系统”的共同待审申请中进行了描述。

5 通信信息的不同部分可能存在不同的 TIS。例如，一通信信息可包括一引用部分，该部分具有与该通信信息的其余部分不同的音调。为了获得高质量的通信信息，大多数文档的 TIS 具有与引用部分相关的 TIS 不同的参数集。为便于说明，仅对与单一通信信息相关的单一 TIS 进行描述，但是本发明可扩展到每个通信的多个 TIS。

10 在图 1 中，当用户请求一网页时所述方法开始进行。用户浏览器对应 TIS 而对网页进行句法分析。一示例性的一般 TIS 可能具有如下结构：

```
<?xml version="1.0" encoding="UTF-8? ">
<TIS>
    < Version >1.0</ Version >
    <SourceLang>en< /SourceLang >
    < MIME-Type >text/rtf< /MIME-Type >
    < Encoding >ISO8859-1</Encoding>
    < Tense >
        < Item1 >formal</Item1>
        < Item2 >business</Item2>
    </Tense>
    < AvailableTranslation >
        < de_DE >http://www.source.com/reference</de_DE >
        < fr_CA >file://lanhost//d:/path/docname</fr_CA >
    < /AvailableTranslation >
    < TranslationMemory >
        < Item1 >TM-1 reference</Item1>
        < Item2 >TM-2 reference</Item2>
    < /TranslationMemory >
    < Service >
        < Engine >special engine xy2</ Engine >
        < PreferredAgency >Worldlingo</PreferredAgency>
```

```
<PreferredTranslator>
  < Item1>
    <Language>de </ Language >
    <Name>Hans Schmidt </Name>
  5      </ Item1>
  < Item2>
    <Language>it </ Language >
    <Name>Bruno Zagani </Name>
  </ Item2>
10     </PreferredTranslator>
</Service>
<Dictionary>
  < Item1>Dictionary1 reference</Item1>
  < Item2> Dictionary 2 reference</Item2>
15     </Dictionary>
<Glossary>
  < Item1> Glossary 1 reference</Item1>
</Glossary>
<DNT>
  < Item1>Microsoft</Item1>
  < Item2>Wordlingo</Item2>
</DNT>
<DNT—List>
  < Item1> http://www.source.com/dnt-list-doc</Item1>
20     </DNT—List>
<Use>
  < Item1>marketing</Item1>
</Use>
<Industry>
  < Item1>engineering</Item1>
</Industry>
<Rulesets>
  <Tagged>
    <Tagged1>
30      <Type>html</Type>
  35
```

```

< Item1>
  <Name>rules</Name>
  < Item1>
    <Expr>if (stillTranslate) then translateContent()</Expr>
5
    </ Item1>
  < /Item1>
  < Item2>
    <Name>base</Name>
    < Item1>
      <Expr>if      (hasAttribut("href"))      then      parse(a
ttribute("href"))</Expr>
      </ Item1>
    < Item2>
    </Tagged1>
15
    </Tagged>
  </Rulesets>
</TIS>

```

实际的标识符随着任意给定的情况而不同，并且可能包括本例中所给出的上述标识符的子集以及未给出的附加标识符。

20 在一般的 XML TIS 中所给出的标识符具有下述功能：

<TIS>...</TIS>
 ——标明 TIS 的开始和结束。
 <Version>...</Version>
 ——表示 TIS 结构的版本
 <SourceLang>...</SourceLang>
 ——标明通信信息语言
 < MIME-Type>...</MIME-Type>
 ——表示 MIME 类型
 <Encoding>...</Encoding>
 ——表示编码
 < Tense>...</Tense>
 ——表示时态，这由机器翻译引擎来读取，其可用以作为提高翻
 译质量的参数；

<Item#>...</Item#>

——确定多个项目的优先次序。项目 1 应用于项目 2 之前。

<AvailableTranslation>...</AvailableTranslation>

——可用的/优选的翻译的列表。例如，一网页可能已经具有一个

5 外语对等网页，可传送该外语对等网页以代替所访问的网页；

<TranslationMemory>...</TranslationMemory>

——指向一翻译存储器的指示字，其中该翻译存储器从一高速缓冲存储器中检索翻译以避免重新翻译已译文档或部分文档；

<Service>...</Service>

10 表示首选诸如特定翻译引擎或特定人工翻译这样的翻译；

<Engine>...</Engine>

——优选的引擎；

<PreferredAgency>...</PreferredAgency>

——用于执行所需翻译的优选机构；

<PreferredTranslator>...</PreferredTranslator>

15 优选的人工翻译者，也许要对应于每种语言；

<Dictionary>...</Dictionary>

——所要使用的一个翻译词典或多个词典；

<Glossary>...</Glossary>

20 所要使用的一个翻译词汇或多个词汇；

<DNT>...</DNT>

——不翻译的词或短语的列表；

<DNT-List>...</DNT-List>

25 指向一文件的指示字，其中该文件包含有不翻译的词或短语的列表；

<Use>...</Use>

——翻译引擎所使用的一上下文标识符，用以提高翻译质量；

<Industry>...</Industry>

——用于提高翻译质量的又一上下文标识符；

30 <Rulesets>...</Rulesets>

——在翻译过程中可使用的规则或指南的列表。

一旦标识了 TIS，浏览器即析取翻译参数并执行相应动作。第一个

动作是依照用户的优选语言来检查源语言标记符。从操作系统设置中、从 cookie 文件即存在于接收计算机或其它可存取位置的优选文件中、或者从适当软件所执行的分析中可获得用户的优选语言。如果优选语言与源语言相匹配，则不必翻译而显示其页面。如果不匹配，则可获得翻译。

TIS 不局限于方便通信信息的无缝翻译。TIS 还可提高非无缝翻译系统中的翻译质量，该非无缝翻译系统例如可以是前面所提到的本发明人的共同待审申请中所描述的一次点击翻译系统。图 2 给出了利用 TIS 的非无缝翻译系统。

当用户接收到诸如电子邮件并且该电子邮件程序显示了该电子邮件时，则开始图 2 所示的处理。如果电子邮件不是接收者的优选语言，则进行诸如单击一次点击翻译组件这样的动作以请求翻译。翻译管理器将电子邮件解析成一 TIS。分析 TIS 的内容并根据包含在 TIS 中的信息来获得翻译。对于电子邮件而言，这通常意味着将翻译参数提供给翻译引擎。

TIS 的优点在于它可将所请求的通信信息重定向到外文对等通信信息。许多公司以多种语言来维持镜像站点。如前面的一般例子所示，TIS 可包含指向这些站点的标识字。

在一个例子中，这通过利用规则以影响镜像页面位置来实现，从而无需为每个 URL 指定本地网页名称。例如，TIS 可包含一个指向服务器中的词典的标识字，该服务器存储适用于该通信的所有本地.html 文件（网页）。TIS 还可包含一个指向一个规则或一组规则的指示字，该规则表示当对通信信息段进行处理时如何处理.html 文件。

例如当将通信信息翻译成日语时，所述规则表示可用 X_japanese 来代替 X 或 X_*（其中*是通配符并且可表示任意的扩展名）。

只要对这段信息进行处理并且对被称为 homepage.html 的网页进行处理，如果 homepage_japanese.html 处于特定的位置，则 TIS 指向用 homepage_japanese.html 来代替 homepage.html 的规则。类似地，TIS 可提供一个指向其中存储有本地图形这样位置的指示字，使得在 tree_homepage_japanese.gif 是可用的情况下，可从特定的位置获得 tree_homepage_japanese.gif 并且其将包含在翻译内以代替

tree_homepage.gif。

如上所述, TIS 不局限于网页。还可按照类似的方式利用公知的 v 卡将 TIS 添加到电子邮件中。下面的例子给出了利用诸如 “text/x-tis” 这样的自定义 MIME 型(也被称为内容型)将 TIS 附加到电子邮件中。

5 Content-Type: text/x-tis; name= “settings.tis”
 Content-Transfer-Encoding: quoted-printable
 Content-Disposition: attachment; filename= “settings.tis”
 <TIS>
 ...
10 </TIS>

将诸如前面的例子所给出的 TIS 嵌入到一电子邮件的单独部分中。电子邮件内的这些单独部分是常规惯例, 其表示给定电子邮件内容的附件。TIS 作为附件的优点就是它不受因特网上传输的影响并且不取决于个别邮件服务器的邮件处理系统。

15 另一个用于实现电子邮件中的 TIS 的实施例是按照下述方式将其添加到自定义的头部:

Received: Laptop (isp.net[192.168.41.217]) by wlm.worlding.com with
SMTP (Microsoft Exchange Internet Mail Service Version #)
id 1BV3QHTA; Fri, 9 Feb 2001 09:29:49 +1000
20 From: “###” <#@worldlingo.com>
 To: “###” <#@worldlingo.com>
 Subject: 专利
 Date: Thu, ## #### ##
 Message-ID: <#####.#@worldlingo.com>
25 MIME-Version: 1.0
 Content-Type: text/plain;
 charset="iso-8859-1"
 Content-Transfer-Encoding: 7bit
 X-Priority: 1 (highest)
30 X-MSMail-Priority: high
 X-Mailer: Microsoft Outlook IMO, Build 9.0.2416 (9.0.2911.0)

X-MimeOLE: Produced By Microsoft MimeOLE V5.50.4133.2400

X-TIS-Version: 1.0

X-TIS-SourceLang: en

X-TIS-Service: Engine= engine1

5 X-TIS-Tense: formal, business

Importance: high

该实施例仅仅给出了先前在一般的 XML 结构例子中所标注的可能的 TIS 字段的几个字段。对于任一特定应用，只有所需字段必须包含在 TIS 中。用公知且合理的缺省值来代替不需要的字段，或简单地省略不需要的字段。

10 TIS 的报头实施例还可适用于文本、RTF、或专用格式的文档。大多数文档包含了专用于规定文档外观的报头信息。可将 TIS 添加到该报头信息上，以便在接收者查看之前可对该文档进行无缝翻译。还可将 TIS 添加到利用 MSWord 或其他专用字处理器所创建的文档的属性对话框中。

15 TIS 还可被包括作为以下 HTML 注解模块的例子所示的 HTML 文档的一部分。

```
<html xmlns:t="urn:schemas-worldingo-com:itis:itis"
      xmlns="http://www.w3.org/TR/REC-htm140">
 20    <head>
        <meta name="X-TIS-Version" content="1.0">
        <title>Reference Document</title>
        <!--><xml>
        <t:TIS>
            <t:SourceLang>en</t:SourceLang>
            <t:Version>1.0</t:Version>
            <t:MIME-Type>text/rtf</t:MIME-Type>
            <t:Encoding>ISO8859-1</t:Encoding>
            <t:Tense>
                <t:Item1>formal </t:Item1>
                <t:Item2>business </t:Item2>
            </t:Tense>
 30
```

```
<t:Industry>
  <t:Item1>engineering</t:Item1>
</t:Industry>
</t:TIS>
5  </xml><!-->
```

另一个解决方案是利用 HYML 文档的 TIS，如以下 HTML 中间标记符这个例子所示。

```
<HTML>
  <HEAD>
10   <META name="X-TIS-Version" content="1.0">
    <META name="X-TIS-SourceLang" content="en">
    <META name="X-TIS-Service" content="Engine|engine1">
    <META name="X-TIS-Tense" content="formal, business">
    <TITLE>Search Results</TITLE>
15   </HEAD>
  <BODY>
...
</BODY>
```

TIS 不局限于文本应用。基础的翻译引擎适用于将语音翻译成文本、将文本翻译成语音、以及将语音翻译成语音。TIS 可通过定义诸如音调、重音、内容、以及字段这样的参数而明显提高这些基础翻译引擎的有用性。

图 3 给出了该 TIS 在一网络环境中的实际实施方案的示意图。用户 1 利用个人计算机 3 上的浏览器来请求或接收诸如网页 2 这样的通信信息。浏览器通过因特网 5 而向网络服务器 4 请求网页 2，并将其显示在个人计算机 3 上。

如果对于该用户来说通信信息是外语，用户 1 可能请求翻译。如上所述，根据本发明人共同待审的申请中所描述的处理而自动出现了本步骤。个人计算机上的浏览器将分析通信信息的 TIS，并利用从 TIS 30 中所获得的参数，通过因特网 5 请求翻译。

如果网络服务器 4 具有通信信息 2 的适当翻译 2a，则将其直接提供给用户 1。如果没有适当的翻译，则将翻译请求以及得自 TIS 的参数

一起传送到翻译管理器 6。翻译管理器 6 从翻译引擎 7 获得翻译 2b。

为便于说明，所示翻译管理器 6 和翻译引擎 7 是独立的。这些功能也可具体实施为运行在一个计算机上的单一应用程序或独立应用程序。如果安装了适当的软件，则翻译功能甚至可在个人计算机 3 上本地执行。
5

可由接收通信信息的应用程序来读取 TIS，但是并不局限于这一实施方案。如果翻译引擎位于接收者的计算机中，或是位于与接收者相连的一网络中时，翻译引擎可直接解释该 TIS。更适合的是，可将网络中的服务器配置成一翻译管理器，在将通信信息传送到接收者之前，
10 该翻译管理器检测 TIS 并对翻译信息的翻译进行管理。翻译管理器也可位于接收者的计算机中。

为了有助于理解本发明，下面对特定情况的 TIS 应用程序进行详细的说明。

大多数现有在线交易发生于西方并以英语设计其网页和文档。然而，增长最快的互联网访问出现在英语并非是第一语言、并且许多可能的潜在客户甚至不会讲英语的地区。为了开发这些潜在客户的市场，网页和文档必须展示为他们的本地语言。大多数人们将不会费力对网页进行翻译，并且当然不会为翻译付费。为了开发这些人的市场，翻译必须是无缝出现的。
15

翻译可以以多种方式出现，而 TIS 有利于所有这些方式。首先，接收者具有位于其计算机中的机器翻译软件。在这种情况下，TIS 提供所有的相关参数，从而导致无缝显示高质量的翻译。其次，接收者与因特网相连，因此 TIS 将网页传送给翻译管理器，该翻译管理器进行必要的翻译并将其无缝显示给接收者。第三，创办人可能已创办了相关语言的镜像站点，在这种情况下 TIS 无缝地将接收者的浏览器导引至该镜像站点。
25

为取得最大效率，包含在 TIS 中的参数必须是相关的并且可被所采用的翻译引擎所理解。因为存在很广范围的翻译引擎，因此实现该要求很困难。然而，发明人已实现了 TIS 包含可扩展的类属参数集。
30

对于机器翻译引擎而言对 TIS 进行解释并将类属参数转换成专用命令，这是相对简单的问题。发明人预计还可产生转换程序，从而使传统的

翻译引擎可解释类属 TIS 参数。

本发明的一个重要优点就是灵活性，它允许用户为特定的通信信息段自定义执行翻译的方法。TIS 提供了指向特定信息、规则、指南以及资源的指示字，而非一般的宽泛的方法，这可使用户获得更好的翻
5 译。

自定义这样的灵活性——且不局限于此——对于因特网尤其重要，这不需从头到尾地读取文档和通信信息段，而是可导航通过。另外，用来进行导航的方法可由每个网络管理员所设计/规定——例如通过其编写自定义的 Java 描述语言。因此一般的方法不能涵盖所有的情况。
10 因此随着用户导航通过整个文档或通信信息段，自定义的灵活性对于所要适当执行的翻译是必需的。

在另一个实施例中，两个（2）或更多个 TIS 参数可彼此影响，从而提供较好的翻译。例如，TIS 可包含一个指向服务器中的字典的指示字，可应用于该通信信息段的所有本地图形均存储在该服务器中。TIS
15 还可包含另外一指示字，其指向一个规则或一组规则，该规则涉及当对通信信息段进行处理时如何对图形进行处理。

参考前述实施例，当将通信信息段翻译成日语时，规则可能会表示用 X_japanese 来代替 X 或 X_*(其中*是通配符并可表示任意的扩展名）。因此，随着通信信息段受到处理且被称为 tree.gif 的图形受到处理，
20 TIS 指向一规则，该规则指示用 tree_japanese.gif 来代替 tree.gif，并且 TIS 提供了一个指向存储有本地图形的位置的指示字，从而当在该存储位置中有 tree_japanese.gif 时，就可获得 tree_japanese.gif 并将其包含在翻译中。

本发明的 TIS 的优点在于一个 TIS 参数影响另一 TIS 参数的能力，
25 这可简化翻译系统的实现、管理及维护以及它所用以执行翻译的资源。这使得用户可对诸如图形这样的本地组分定义命名规则。例如，将
“_Language”添加到每个目标语言的图形名称上，其中在上面的例子中“_Language”是“_Japanese”。可如上所述地编写一个通用规则，而不是为每个单独的图形编写一特定规则。

30 一 TIS 参数对另一 TIS 参数的影响可用于文档和/或不同文件类型内的更多组分。

因特网决定出自一组织机构的大约什么信息将被翻译。如果用户想要翻译，则他们可容易地从多种便宜或免费的在线翻译网站中获得。组织机构的挑战是要确保从这些在线翻译站点所获得的翻译以优选的方式描绘该组织机构和/或其产品及服务。TIS 为组织机构提供了一渠道以公开其翻译优点及资源，从而使用户能够获得更准确的、更能描述出组织机构所寻求信息的译文。

尽管说明书的目的在于对本发明的实施例进行详细的描述，但是本发明并不局限于任意一种可替换特征的特定组合。

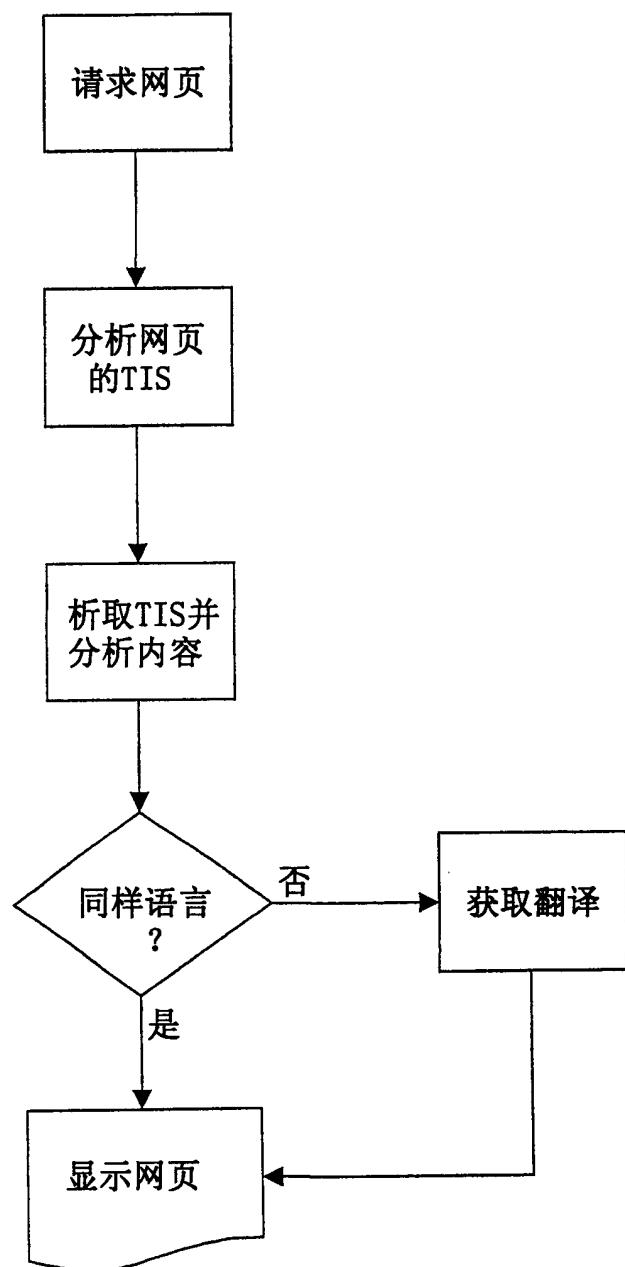


图1

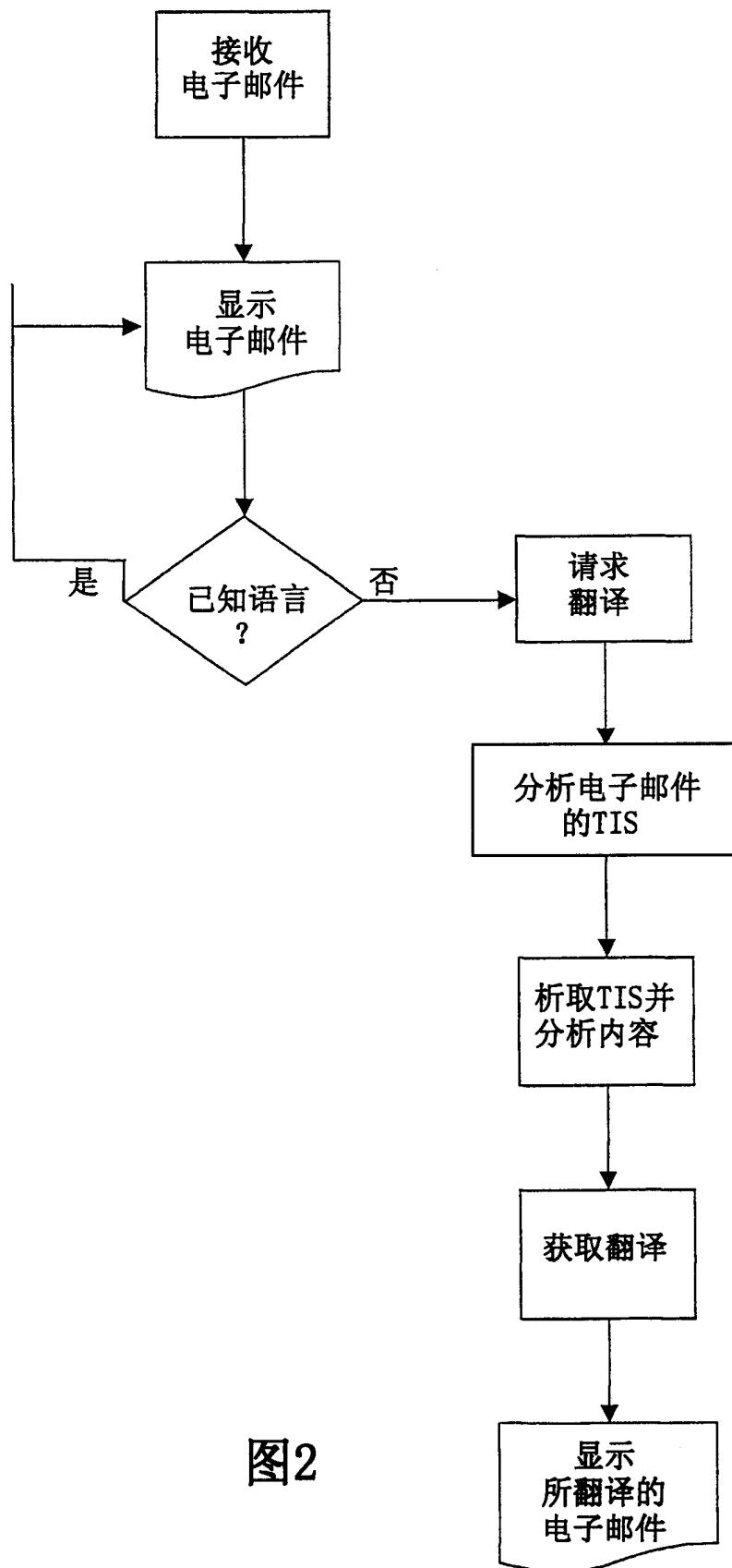


图2

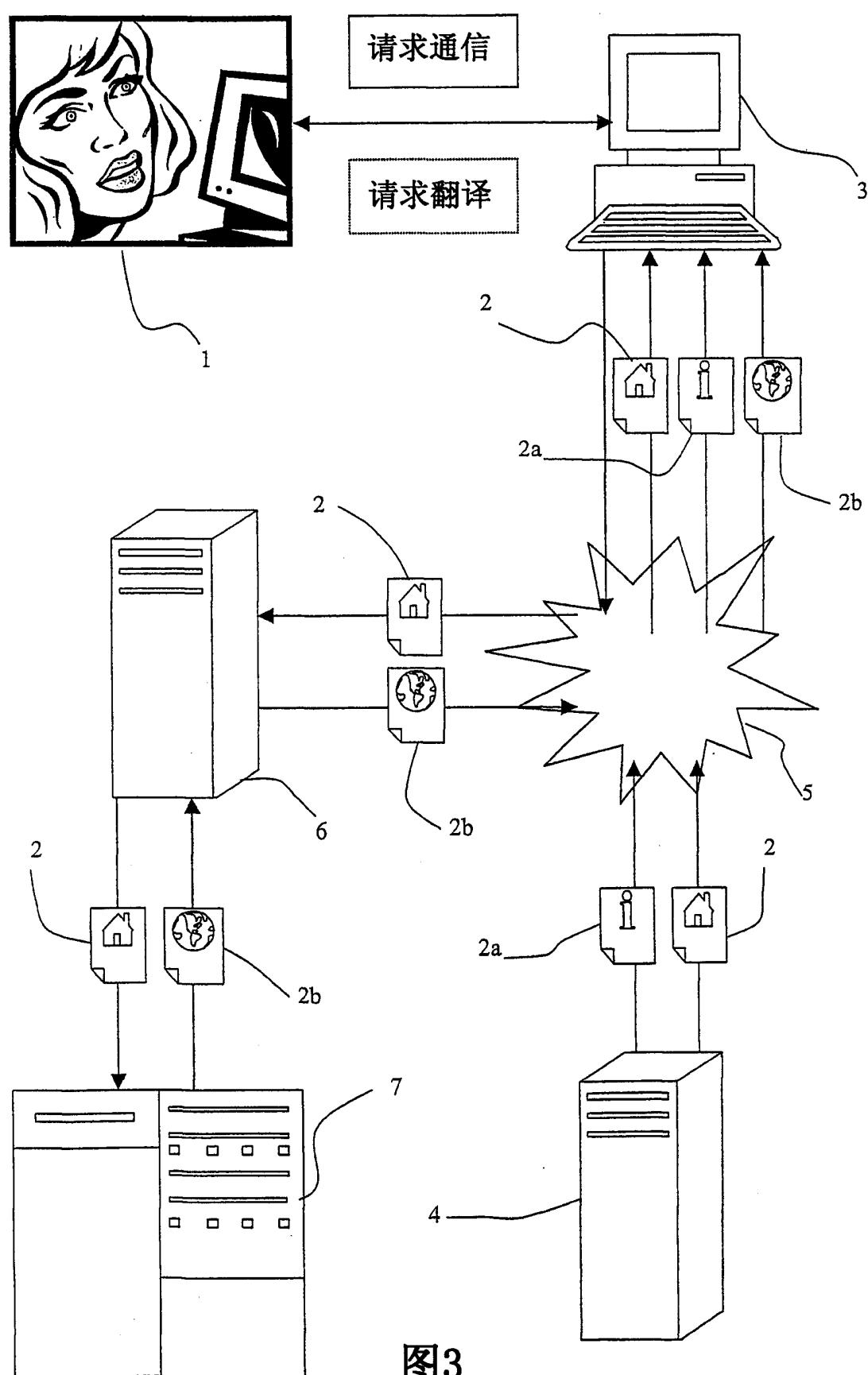


图3