发明名称
一种用于移动通讯设备终端的网页内容识别管理系统

摘要
本发明涉及移动终端领域，特别是一种用于移动通讯设备终端的网页内容识别管理系统。本发明提供了一种用于移动通讯设备终端的网页内容识别管理系统，该系统安装在移动通讯设备终端，系统包括用于复制保存网页内容到移动通讯设备终端的复制模块，复制保存移动通讯设备终端的网页内容识别模块，以及用于获取网页内容的网页内容识别模块。复制保存移动通讯设备终端的网页内容识别模块，能够复制保存当前正在浏览的网页的所有内容，并将其存储在移动通讯设备终端中。本发明提供的方案，通过复制和存储网页内容，使用户能够更方便地获取所需的信息。
1. 一种用于移动通讯设备终端的网页内容识别管理系统，该系统安装在移动通讯设备终端，系统包括用于复刻保存网页内容到移动通讯设备终端的复制模块，其特征在于，所述复制模块包括：
   (11) 复制保存移动通讯设备终端光标选择内容的动态光标模块；
   (12) 复制保存移动通讯设备终端当前屏幕显示的所有内容的窗口光标模块；
   (13) 复制保存当前正在浏览的网页的所有内容的网页光标模块。
   所述的系统还包括对复制模块保存的网页内容中的文字部分进行分词处理的分词模块；对分词模块进行分词处理后的词进行搜索的搜索模块。
   所述搜索模块具体用于：列出分词模块进行分词处理后的词组，在由用户选择确定需要进行网络搜索的词组后，以用户确定的词组为关键字，通过网络搜索引擎进行搜索，向用户返回网络搜索结果。

2. 根据权利要求1所述的网页内容识别管理系统，其特征在于，所述的动态光标模块采用如下步骤复制保存网页内容：
   (21) 当用户使用光标移动选择内容时，保存当前页面上的可选内容；
   (22) 当用户使用光标，以手动方式选择内容，则保存用户选择的内容。

3. 根据权利要求1所述的网页内容识别管理系统，其特征在于，所述的窗口光标模块采用如下步骤复制保存网页内容：
   (31) 确定当前屏幕显示中第一行内容在整个网页中的坐标，确定为初始坐标，包括横向坐标和纵向坐标；
   (32) 向移动通讯设备终端查询移动通讯设备终端实际显示尺寸的长度和宽度；
   (33) 以初始坐标和移动通讯设备终端实际显示的长度和宽度作为矩形，矩形的左上角为初始坐标，矩形的长度为移动通讯设备终端屏幕的长度，矩形的宽度为移动通讯设备终端屏幕的宽度；
   (34) 保存该矩形内的网页内容。

4. 根据权利要求1所述的网页内容识别管理系统，其特征在于，所述的分词模块采用基于多层隐马模型的汉语词法分析系统进行分词处理。

5. 根据权利要求1所述的网页内容识别管理系统，其特征在于，所述的系统还包括用于发送所述复制模块保存的网页内容的发送模块。

6. 根据权利要求5所述的网页内容识别管理系统，其特征在于，所述的发送模块发送网页内容的步骤如下：
   (91) 发送模块获取复制模块保存的网页内容；
   (92) 判断网页内容中是否全部由文字组成；
   (93) 假如网页内容全部由文字组成，则调用移动通讯设备终端的短信息服务功能，把网页内容作为待发送的短信息服务内容；
   (94) 用户对待发送的短信息服务内容进行编辑确认后，通过短信息服务方式发送；
   (95) 假如网页内容中包括非文字部分，则只保留移动通讯设备终端能通过多媒体信息服务功能发送的网页内容；
   (96) 调用移动通讯设备终端的多媒体信息服务功能，保留的网页内容作为待发送的多媒体信息服务内容；
（97）用户对待发送的多媒体信息服务内容进行编辑确认后，通过多媒体信息服务方式发送。

7. 根据权利要求6所述的网页内容识别管理系统，其特征在于，所述的能通过多媒体信息服务功能发送的网页内容为文字或者图片或者音乐或者flash。
一种用于移动通讯设备终端的网页内容识别管理系统

技术领域
[0001] 本发明涉及移动终端领域，特别是一种用于移动通讯设备终端的网页内容识别管理系统。

背景技术
[0002] 移动通讯设备终端与普通个人电脑终端不同，普通个人电脑的应用方式和操作方式极为简便，但移动通讯设备终端输入操作较为繁琐。而且移动通讯设备终端的软件往往会缺少一些常用的操作功能。比如最常用的复制粘贴。这导致了移动通讯设备终端用户在上网浏览时，无法实现一些快捷操作，比如选中页面内容进行快速搜索，比如复制选中页面内容，通过短信息服务方式发给朋友。

发明内容
[0003] 本发明提供一种用于移动通讯设备终端的网页内容识别管理系统，以解决现有技术中，移动通讯设备终端用户在上网浏览时无法对网页进行复制以及进行快捷操作的技术问题。
[0004] 本发明提供了一种用于移动通讯设备终端的网页内容识别管理系统，该系统安装在移动通讯设备终端，系统包括用于复制保存网页内容到移动通讯设备终端的复制模块，复制模块包括：
[0005] （11）复制保存移动通讯设备终端文档选择部分的内容的动态光标模块；
[0006] （12）复制保存移动通讯设备终端屏幕显示的所有内容的窗口光标模块；
[0007] （13）复制保存当前正在浏览的网页的所有内容的网页光标模块。
[0008] 采用上述不同的复制方法，能满足用户的不同需求，对用户正在浏览的网页内容进行有选择性的复制保存。
[0009] 作为一种优选方案，动态光标模块采用如下步骤复制保存网页内容：
[0010] （21）当用户没有使用光标手动选择内容时，保存当前页面中的可选内容；
[0011] （22）当用户使用光标，以手动方式选择内容时，则保持用户选择的内容。
[0012] 窗口光标模块采用如下步骤复制保存网页内容：
[0013] （31）确定当前页面显示中第一行内容在整个网页中的坐标，确定为初始坐标，包括横向坐标和纵向坐标；
[0014] （32）向移动通讯设备终端查询移动通讯设备终端实际显示尺寸的长度和宽度；
[0015] （33）以初始坐标和移动通讯设备终端实际显示的长度和宽度作一个矩形，矩形的左上角为初始坐标，矩形的宽度为移动通讯设备终端屏幕的长度，矩形的宽度为移动通讯设备终端屏幕的宽度；
[0016] （34）保存该矩形内的网页内容。
[0017] 作为更进一步的优选方案，本网页内容识别管理系统还包括对复制模块保存的网页内容中的文字部分进行分词处理的分词模块。
[0018] 特别是，分词模块采用基于多层隐马模型的汉语词法分析系统进行分词处理。
[0019] 通过对文字采用分词处理，则能把其中的分词部分分解成具体的有意义的单字或词组，以供后面的搜索分析使用。
[0020] 作为更进一步的优选方案，系统还包括对分词模块进行分词处理后的词进行搜索的搜索模块，其具体步骤如下：
[0021] (71)列出分词模块进行分词处理后的词组；
[0022] (72)由用户选择确定需要进行网络搜索的词组；
[0023] (73)以用户确定的词组为关键字，通过网络搜索引擎进行搜索；
[0024] (74)返回网络搜索结果。
[0025] 其中的词组是由分词模块进行分词得到的，由于通过了分词技术，因此用户能把整段的字分解成单字或者词组，则能以单字或词组作为搜索的关键字，获取用户所感兴趣的信息。假如不采用分词技术，则只能把整段的字作为搜索的关键字，就不能准确地得到用户所感兴趣的信息。
[0026] 作为另外一种优选方案，系统还包括用于发送所述复制模块保存的网页内容的发送模块，其具体步骤如下：
[0027] (91)发送模块获取复制模块保存的网页内容；
[0028] (92)判断网页内容中是否全部由文字组成；
[0029] (93)假如网页内容全部由文字组成，则调用移动通讯设备终端的短信息服务功能，把网页内容作为待发送的短信息服务内容；
[0030] (94)用户对待发送的短信息服务内容进行编辑确认后，通过短信息服务方式发送；
[0031] (95)假如网页内容中包括非文字部分，则只保留移动通讯设备终端能通过多媒体信息业务服务功能发送的网页内容；
[0032] (96)调用移动通讯设备终端的多媒体信息服务功能，保留的网页内容作为待发送的多媒体信息服务内容；
[0033] (97)用户对待发送的多媒体信息服务内容进行编辑确认后，通过多媒体信息服务方式发送。
[0034] 通过上述的发送方式，用户能快捷地把正在浏览的网页内容通过短信息服务和多媒体信息服务发送给朋友分享，使通信更加方便。
[0035] 其中，能通过多媒体信息服务功能发送的网页内容为文字或者图片或者音乐或者flash。
[0036] 本发明提供的方案，通过对整个页面的复制功能以及后续的处理方法，提高网页浏览相关软件的实际应用能力，为用户带来良好的使用体验，使用户更方便地对其所感兴趣的内容进行搜索和分享，提高了网络互动性。

附图说明
[0037] 图 1 为光标定位方法。
[0038] 图 2 为本实施例的分词搜索示意图。
[0039] 图 3 为本实施例通过短信息服务发送的示意图。
具体实施方式

[0040] 下面结合附图和具体实施方式对本发明作进一步详细的说明。

[0041] 本发明提供了一种用于移动通讯设备终端的网页内容识别管理系统，该系统安装在移动通讯设备终端，系统包括用于复制保存网页内容到移动通讯设备终端的复制模块，复制模块包括：

[0042] (11) 复制保存移动通讯设备终端光标选择部分的内容的动态光标模块；
[0043] (12) 复制保存移动通讯设备终端屏幕显示的所有内容的窗口光标模块；
[0044] (13) 复制保存当前正在浏览的网页的所有内容的网页光标模块。
[0045] 通过上述的复制方法，本发明能对用户正在浏览的网页内容进行有选择性的复制保存。

[0046] 上述三种光标模块具体说明如下：

[0047] A) 动态光标模块，如图 1 (a) 所示，主要是指页面上可以被感知的光标(超链接被选中光标)，这种光标可分为两部分：

[0048] A1) 默认光标，如图可选中页面上可选资源那些光标；
[0049] A2) 手动光标，以手动划选页面内容的光标(类似电脑终端上鼠标左键选中)。

[0050] B) 窗口光标模块，如图 1 (b) 所示，以当前浏览窗口为标准，当前窗口内容都是被光标选中的内容，具体步骤如下：

[0051] B1) 确定当前屏幕显示中第一行内容在整个网页中的坐标，确定为初始坐标，包括横向坐标和纵向坐标，即图中坐标(x, y)；
[0052] B2) 向移动通讯设备终端查询移动通讯设备终端屏幕的长度和宽度；
[0053] B3) 以初始坐标和移动通讯设备终端屏幕的长度和宽度作一个矩形，矩形的左上角为初始坐标(x, y)，矩形的长度为移动通讯设备终端屏幕的宽度，矩形的宽度为移动通讯设备终端屏幕的宽度；

[0054] B4) 保存该矩形内的网页内容。

[0055] C) 画面光标模块，如图 1 (c) 所示，以当前浏览页面为标准，当前页面上内容都是被光标选中的内容。

[0056] 程序根据以上三种光标模块识别并提取页面上对应的内容；

[0057] 程序根据识别出的内容提供一系列用户可能需要进行的功能操作；

[0058] a) 复制。将识别出的内容复制到缓存模块(可以为暂时的系统内存，也可以为长期的本地存储)，以供使用这些数据的话，可通过“粘贴”操作进行调用；

[0059] b) 搜索。在页面内容识别的基础上，提供搜索功能；将选中的文字内容使用“分词技术”进行划分，得出最适合的分词组合，用户通过选择并“确定”就能将内容提交到搜索引擎进行搜索。

[0060] 分词技术采用比较流行的基于多层隐马模型的汉语词法分析系统进行分词处理。

[0061] 如图 2 所示，对于短句“暑假去哪里度假好？推荐一下，通过分词技术分解出暑假、度假、推荐三个词组，这三个词组分别为例，实际的分词结果可能与此不同)，使用这三个词组作为关键字在流行的搜索引擎如百度、谷歌、易搜中进行搜索，获得搜索结果返回给用户。
在页面内容识别提取的基础上，调用移动通讯设备标准上已有的通信功能（包括但不限于短信息服务、多媒体信息服务），将选中内容发送给朋友。

如图3所示，对于短句“暑假去哪里度假好？推荐一下”，调用系统的发送短信息服务功能，把短句“暑假去哪里度假好？推荐一下”作为短信息服务的内容部分进行发送。

上述仅为细节举例说明，实际范围不限于此。
坐标 (x, y)

图 1