

# 基于上下文的个人信息管理研究

刘芳 李春旺

**【摘要】**文章在介绍个人信息管理及上下文的概念和特点的基础上,对个人信息管理的上下文情景模式、基础架构、本体模型及相关检索机制进行总结和概括,并以 Feldspar、Hystack、Gnowsis 及 Semex 为例对基于上下文的个人信息管理工具的发展现状做了简单说明。

**【关键词】**信息关联 上下文 个人信息管理 检索

**Abstract:** After introducing the definition, type and feature of context and personal information management, this paper emphatically analyzes and summarizes the mode, architecture and related algorithm of context-based personal information management, and takes Feldspar, Hystack, Gnowsis and Semex as examples simply to explain the development of PIM tools.

**Key words:** information associated context-based personal information management retrieval

## 1 个人信息管理介绍

从 Vannevar Bush 在 1945 发表的论文“*As We May Think*”<sup>[1]</sup>中第一次提出“扩展存储器 (Memory-Extender)”设想开始,到 2005 年起至今每年一届的个人信息管理专题讨论会<sup>[2]</sup>,如何对个人信息资源归纳整理便于再次使用一直是备受专家关注的问题。华盛顿大学发起的 Keep Found Things Found 项目<sup>[3]</sup>中,将个人信息管理定义为学习和模仿个人获取(或创造)、存储、组织、保持、检索、使用和传播信息,以实现生活目标和社会角色的行为方式。个人信息管理通常侧重于对个人收集或创建的、日后再次使用信息的组织和保持,其实质可以概括为在信息和用户的需求间建立映射关系,以实现信息的重新发现。个人信息本身具有的特点会为管理带来一定困难,主要表现在以下几方面。首先,以不同类型和格式存在的大量异构信息片段需要抽取信息、建立统一格式的索引<sup>[4]</sup>;其次,不同于网页间的点击链接及基于内容的检索方式,个人信息的外部属性间存在多种上下文关联,这些关联对于信息发现起着重要作用,Google、Baidu 等桌面搜索工具不能取得较好的检索效果<sup>[5]</sup>;最后,随着 Web2.0 应用的普及,个人信息将越来越多地渗入到社会网络中,与组织信息管理 (GIM) 和社会语义网 (FOAF、SWAP) 相融合,广义上下文情景的个人信息管理受到重视<sup>[6]</sup>。

## 2 个人信息的上下文情景

### 2.1 上下文情景的定义及特点

上下文情景是指个体在对某个给定目标进行推理的过程中用到的事实和推理规则的集合<sup>[7]</sup>,包括环境状态、资源信息、用户需求等影响输入输出的一些属性及属性的取值。从理论上说,上下文情景与用户心中个性化的“情景”类似,它由特殊的事件及环境引起,与要完成的任务目标具有松散的关系,当任务越来越清晰时,依据该观点由计算机自动生成上下文情景也更加具体。上下文情景是暗含的、潜在的,当偏离预期的事件发生时,上下文情景也会随之波动<sup>[8]</sup>。上下文信息能否在系统的运行和交互过程中发挥作用,主要取决于以下两个方面<sup>[9]</sup>:(1) 如何从各种信息关联中形成有用的上下文信息。上下文情景的一个最突出特点是种类繁多,个人信息空间中的诸多因素都可能对系统产生影响。(2) 上下文信息本身和它们之间的相互关系往往随着系统的使用而不断变化,是否记忆这些变化,并把这些变化表现出来可能需要借助语义关系结构。检索中常用的上下文关系包括:面向语料库的上下文关系(如词语聚类、布尔分解、潜在语义索引)、面向个人历史的上下文关系(如用户的浏览日志、历史查询)及面向查询特征的上下文关系(如摘要、标题、关键字)等。但是对于个人信息,应充分考虑其信息环境中包含的特殊的上下文关系。

## 2.2 个人信息的上下文情景类型

通过对目前存在的个人信息管理工具的比较和分析,可以总结出以下几种常用的上下文情景<sup>[10]</sup>:

(1) 基于层级的上下文情景。该类上下文关系主要指本地或远程创建的文件树。文件系统和邮件客户端等都支持用户创建文件夹,子文件夹按照基于主题或类型的方式对自己的信息资源进行分类整理,是最常用的信息关联方式。例如,丽江拍摄的照片可能存在于“丽江”、“云南”、“旅游”这样的文件层级结构中。这类上下文情景符合用户的认知方式,可帮助用户缩小检索空间并消除词语歧义,但要求用户具有清晰的认知及信息组织能力,不适用于非结构化信息及与多个异构信息片段相关的基于任务的检索。

(2) 基于标签的平级上下文情景。为信息项添加标签,利用标签间的关系及多重标签形成的文档分类灵活地对信息进行管理。与层级上下文相比,标签采用了一种轻量级的机制为信息建立多种关联,减轻了用户的认知负担。目前有自动和人工添加标签两种形式。自动标签由系统根据文档的读取路径和文档间的交互信息形成,可用性较高;人工标签种类较丰富,如在 Delicious、SNS 社区内创建的网络标签、浏览器内创建的网页标签、用户添加的信息注释等,但是由于缺乏规范性及过于庞大的数量可能会给系统带来负担。

(3) 基于线性的上下文情景。将信息条目按照某一规则排序,只要用户知道查询信息项的某一属性,就可以生成基于该属性的全部信息的逻辑列表,最典型的是时间序列。线性上下文情景不需要用户进行额外的操作,减轻了用户负担。早期的个人信息管理大部分是基于线性上下文情景的,如 Lifestreams 将所有信息按照时间顺序进行排序,Stuff I've Seen 能够基于时间、作者、文件类型等多条线索进行检索。但是线性索引过于单一,抽取的字段有限并且只停留在信息的外部属性上,没有充分展现个人信息网状的结构特点,当信息量庞大、异构性强时不适于使用。

(4) 基于用户信息操作行为的上下文情景。支持这类观点的学者认为,跟踪并记录用户在使用和存储个人信息过程中的操作行为、操作的信息项和时间,将这些数据添加到信息的时间索引中,可以使用户明确发现关注的重点以及信息的价值点,这是基于个性化对信息进行组织标引的重要途径。相关的信息操作包括:应用程序之间的切换、文件重命名(Rename)、另存为(Save as)、移动(Move)、复制(Copy)和粘贴(Paste)等。该类信息通常被用于任务的发现。完成一个任务通常需要读取和操作多个信息项,例如,写论文时用户会查询相关的介绍、参考文献、网页及同事发来的邮件,并且会在多个信息应用窗口间切换。用户的这些操作行为在某种程度上说明这些信息具有一定的关联性,系统利用基于算法的逻辑推理,将任务相关的信息按照其重要程度展示给用户。该类上下文信息通过对系统及应用程序的监听获取,数据量大冗余度高,需要有效算法的支持。

(5) 基于应用环境的上下文情景。当信息下载到本地后失去了原来丰富的上下文环境,这里主要是指从邮件系统及网络下载的信息。例如,已发送的邮件中含有用户需要解决的问题,而回复邮件的附件里有对这个问题的解释,但是当附件下载到本地后失去与邮件发信人、主题、正文等相关信息的联系。再如用户从网络数据库上下载多篇论文,对于重要文献还会下载它的引用文献,当这些文献下载到用户的个人信息空间时就失去了原来的引用和被引用关系。邮件系统及网络作为个人信息的主要来源经常被作为一个专门的问题进行研究。目前最常用的解决办法是将信息相关项作为元数据抽取出来与信息项建立信息关联,通过推理机制保存信息环境。

(6) 基于网状的上下文情景。基于语义的网状结构更加符合个人信息的组织特点,是近些年来领域内关注的热点及发展方向。一方面,语义网络具有很强的灵活性,通过资源描述框架(RDF)对信息进行组织,并提供唯一标识,允许用户在不同类型的信息间建立关联,还可以利用 RDF 良好的扩展性为文件添加注释和标签;另一方面,由于语义网络结构性较差,无法将个人信息形成清晰的体系或提供有效的信息导向,用户很容易在网络中“迷失”,用户接口、存取控制、语义协同性、推理机制等方面问题需要进一步完善。

## 3 系统框架结构及相关技术

### 3.1 系统框架结构<sup>[11]</sup>

在个人信息管理系统的实际应用中,上下文情景通过抽取信息的内部和外部属性作为元数据并在元数据间建立关联实现。新近开发的个人信息管理工具大部分以语义为基础创建本体对元数据进行描述,将多种类型的上下文情景融合在一起以满足用户的不同检索需求。图 1 是对个人信息管理系统基本框架结构的描述。

从功能层面上看,信息监控系统、服务器和搜索引擎是个人信息管理系统的主要组成部分。信息监控系统负责监听用户的信息增、删、改行为,并通知服务器实时调整存储的数据;服务器负责调度、索引信息,并将检测到的信息发送到相应的信息处理模块中;搜索引擎作为人机交互界面,支持基于内容及语义的检索方式,通过一

定的检索机制从服务器内找到相关信息返回给用户。

为了实现检索任务,系统需要4个模块的支持:(1)元数据抽取。由后台的元数据抽取器统一调度,将信息抽取任务按照类型和内容分配给不同的过滤元件。如利用PDF、DOC等内容过滤元件抽取信息内容,类似于Email内的发送人,题目及文档存储路径等信息则通过相应的元数据过滤元件抽取。(2)元数据富集。该模块又由三部分组成,包括对象识别、算法及推理机制和附件-文档类关系链接。这部分为抽取到的异构信息提供本体模型,允许用户通过对象识别功能为信息项定义唯一标识和添加注释,利用算法和推理机制为属性关系确定权重及权重转移条件,还可将信息划分到不同的任务中,当系统现存的情景关联不能满足需要时支持用户自定义。(3)RDF存储。负责存储信息内容、属性元数据及信息关联。(4)基于关联的信息发现。根据用户的检索要求,依据元数据富集提供的相关过滤条件,遍历RDF仓库,抽取所有相关的信息按照相关程度排序,还可以将信息间的关联可视化给用户。通过上述分析可以看出个人信息管理工具的一个重要特征就是元数据的创建和存储。由于这些元数据的使用和获得是由不同的模块实现的,因此需要一个定义良好的资源描述框架对每种类型的信息和他们之间存在的关联进行描述。

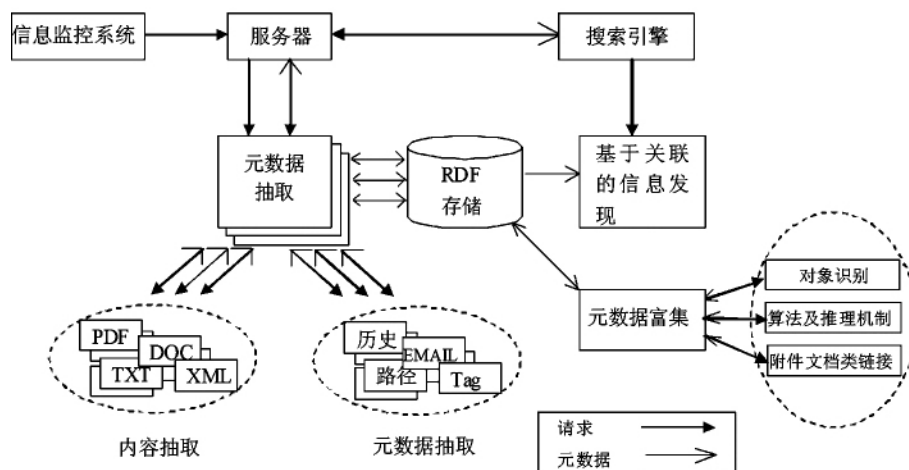


图1 个人信息管理工具系统框架结构

### 3.2 个人信息相关本体模型

Hystack、EPOS (Evolving Personal to Organizational Memories) 等项目先后提出了不同的本体模型规范语义关系,希望能够创建一个通用的、完善的个人信息本体模型,并且能够与社会语义网络建立链接,将个人上下文情境扩大到更广阔的范围,其中以Nepomuk项目的个人信息本体模型(PIMO)应用最为普遍。它将管理范围限定在用户关注的或者工作及私人需要的信息,并将这些信息组建成个人知识空间。该信息空间与用户读取信息的路径与方式无关,并独立于信息的来源、格式及作者,对信息的管理更加人性化,使用户能更加专注于完成他们的任务而不是强调如何找到这些信息。需要指出的是,Nepomuk开发的表示语言(NRL, Nepomuk Representation Language)不同于标准的RDFS,它将三元组存储扩展到四元组对资源进行描述,第四个参数是已定义(存储)的用语句描述的graph(可能为空),用于对描述语句进行分组,如添加Tag标签等。

目前语义桌面内存在的本体模型可以分为三层<sup>[12]</sup>:最底层即基础层是信息本体(NIE),用一种类似于Mimetype的标准,将语义桌面中经常出现的如网页、邮件、多媒体等异构信息依照本体模型进行描述,抽取必要属性及属性关联存储到数据表中,保存信息上下文情境;第二层是个人信息本体模型(PIMO),该模型创建了核心类对人、事物、资源、概念等进行描述并建立基本实体,然后通过PIMO的域本体对关系和类进行扩展,使用户将自己思维方式结构化表示出来,并通过匹配算法和域本体与他人的模型进行交互;第三层是用户分享本体模型,如项目组成员共享同一项目本体、拥有共同爱好的用户创建兴趣组本体等。除去三层基本结构还存在一些其他的本体形式,如基于工作任务的本体、基于用户当前工作情景的本体、社会情景本体等。考虑到一个检索可能会涉及到两个或多个本体模型,需要建立相关的推理机制选择本体模型并给予不同的权重。

### 3.3 相关算法及检索机制

在个人信息管理工具中,上下文情景是作为基于内容的个人信息检索的补充出现的。它的作用主要体现在两

个方面:

(1) 优化内容及语义检索。基于内容及语义的个人信息检索往往将大量相关度较低的信息放在检索结果的 Topn 位置上, 上下文关系可以在检索的过程中发挥作用。由于大量的关联和属性都会影响到信息的排列, 需要相关算法及机制的协同工作, 采用统一方式阐明信息间的关联, 并且利用这种关联进行信息排序。Beagle++ 引入了权重转移架构图的概念提出了 ObjectRank 算法, 为本体模型中的实体和关系添加权重值, 并且在关系网络中延伸。例如某个人对用户来说是非常重要的, 那么他发给用户的所有邮件权重值会比较高, 与邮件相关的其他信息权重也会相对高一些, 重要性会传递下去并依次递减, 由此可以通过公式计算出每个信息的权重。还有些学者认为<sup>[13]</sup> 点击率高的信息资源重要程度越高, 在排序时优先考虑用户经常浏览的信息。

(2) 实现导航式检索。这是用户的信息检索活动中常用的检索方式, 经常与内容检索配合使用。这种检索方式考虑了心理学中人类记忆方式的因素, 研究者认为用户可能无法记住某个信息在电脑中存放的确切位置及相关关键字, 但是记得与该信息相关的其他信息的某些属性特征, 以间接信息为线索依据用户记忆搜索到该信息。这种检索方式较为灵活, 可以根据单个上下文关系缩小信息检索范围, 可以利用两个或多个上下文条件对信息进行多重过滤, 还可以通过关键词找到某个相关信息, 并以该信息项作为检索切入点, 在与该信息相关的所有信息项中找到目标信息。

#### 4 个人信息管理工具介绍

本节以 Feldspar、Hystack、Gnowsis 和 Semex 为例, 对基于上下文关联的个人信息管理工具进行分类介绍。

##### 4.1 导航式个人信息管理工具

Feldspar<sup>[14]</sup> 是卡内基梅隆大学计算机科学人机交互研究所于 2008 年提出的个人信息管理系统, 正处在试验和进一步改进的研发阶段。Feldspar 系统是导航式检索的典型应用, 与 Richard Wurman 提出的 LATCH 理论<sup>[15]</sup> 类似, 通过界定与信息相关的人 (who)、范围 (where)、时间 (when)、类型 (what) 以及其他信息的关系确定信息的准确位置。

该系统基于 Google 提供的 DeskTop API 抽取信息, 并以 “E-mail”、“Folder”、“File”、“WebPage”、“Event” (Outlook 中的日程表)、“Person”、“Date” 7 种类型抽取元数据创建索引, 最后使用 QuickGraph 2.0 在 7 类信息间建立两两关系图表。系统为每一个关系对提供结果生成器, 用于在查询中生成中间检索结果。例如关系对 A-B, 即从 A 序列中返回与 B 序列有关的信息, 其算法机制可以通过一个例子来说明, 寻找 “含有 Spence 发来邮件中附件的文件夹”。在这个例子中需要 3 个结果生成器, 即 folders - files、files - emails 和 emails - persons (Spence)。首先将该查询转化成一个关系序列 folders - files - emails - Spence, 算法从最后一个关系对开始处理, 找出在关系图表中所有与 Spence 有关的信息 email 类信息, 包括发送或从 Spence 处接收到的, 并将所有的 email 项整理到列表中作为结果; 接着将这些结果作为第二次关联的输入值, 生成与这些 email 具有关系的文件列表 (文件可以是任何一个 email 的附件); 最后将文件列表作为第三次的输入。通过三步查询就可以找到含有所有与 Spence 往来邮件中附件的文件夹列表。

这种链式检索机制在生成检索结果时是非常有效率的, 最重要的是它允许添加任何其他类型的数据, 并且在查询时不需要改变算法本身, 只需要为新的数据类型提供新的结果生成器。系统没有使用复杂的语义结构, 消除了传统工具添加上下文信息的局限性, 并提出了一种新的检索方式, 但是目前提供的信息类型和关系较单一。

##### 4.2 基于语义的个人信息管理工具

(1) Hystack<sup>[16]</sup> 是麻省理工学院于 1997 年发起的研究项目, 致力于研发与个人信息管理和语义网络相关的应用程序。其中最突出的研究成果是 Hystack 客户端 (a research personal information manager), 在该开源程序中首次使用了桌面语义技术, 并基于 RDF 动态感知语言开发。为了满足个人对信息组织整理的不同需求, 系统支持用户根据自己的喜好和需要自由地处理信息。具体功能特性表现在: 提供了一套应用性很强的基于语义的数据关联模型; 允许用户根据自己的习惯确定需要被保存、查看和检索信息项, 改变信息检索和导向时展示关系和属性的方式, 并将特定任务的相关信息应用整合入同一工作区; 支持分布式服务, 用户可以进行信息交流和共享。但是由于给予用户过大的权限, 在一定程度上给数据库带来了巨大负担。

(2) Gnowsis<sup>[17]</sup> 是 EPOS 项目的一个研究成果, 是欧盟框架计划支持的 Nepomuk 项目的一部分, 使用了 Nepomuk 创建的个人信息本体模型。该项目始于 2003 年, 由联邦德国教育、科学、研究和技术部支持主办。其主要宗旨是支持用户完成知识密集型的工作, 将个人信息空间和工作环境有效结合起来, 建立语义网络帮助知识管

理和发现。管理内容包括用户的文本操作(如对文档的阅读、检索、创建、存档和分类等),在团队中与同事的交流活动和项目工作的执行情况。Gnowsiss充分利用Web2.0的环境,展现了良好的延伸性及集成性。系统采用面向服务的架构,通过XML-RPC等标准和远程应用程序建立链接,将博客、社区、wiki等网络中的个人信息传送到本地进行处理、存储及整合;使用网络标签对信息进行标注;以语义wikiKaukolu作为前端管理界面,为PIMO内的每一个实体创建wiki页面,利用wiki的plugin功能,协同工作平台及对非结构化数据的良好兼容性。

#### 4.3 调用外部API的个人信息管理工具

Semex<sup>[18]</sup>系统是华盛顿大学计算机科学与工程学院于2004年开发的开源工具。该系统侧重于对Latex、Bibtex、PDF、Word等发表的文献资源创建关联和重新发现。在Semex中,不仅可以检索Person、Publication和Message等信息项,还可以对AuthoredBy、Cites和AttachedTo等关系进行检索。由于信息的异构性并且需要在多种资源间建立关联,能够找到关于某个人(或会议、刊物)的所有资料是非常困难的,这需要加强数据之间的协调性。系统通过调用CiteSeer提供的API与外部数据库建立关联,为用户下载到本地的文献资源保存引用和被引关系;具有强大的数据处理能力,能够将文献中的作者、所属机构以及参考文献的题目、作者、发表刊物、时间等信息作为元数据抽取出来,并有效消除大量元数据冲突及冗余信息,为用户形成清晰的基于研究机构(团体)、会议、人、发表文章、时间、刊物及联系方式的文档参考结构图,帮助用户整理知识结构并发现新的研究方向。

#### 5 结束语

将来,用户的个人信息会融合到一个更大的范围内与他人分享。在这种环境下个人信息不再局限于本地,而是走向一个轻量级的、协同工作的环境。例如,项目组的工作人员会分享同一个日程表、邮件系统、资源平台,这不仅需要将相关应用集中到一个页面上进行管理,还希望分布在不同应用程序中的相关信息片段能够不局限于信息载体而是根据内部逻辑组织起来,并能够同步读取。此外,实现个人信息管理与图书馆服务相结合。将个人在电子图书馆内的浏览历史、电子文献下载、咨询服务、论文发表引用、网络信息及本地资源关联起来对个人信息的上下文环境进行延伸,为知识工作者的科学研究和创新创造条件。Semex通过调用citeseer数据库进行了一定尝试,将中国科学院文献情报中心的E划通系统计划融入个人信息管理工具,为用户提供更加个性化的知识服务。相信随着个人信息管理技术的不断成熟,语义网络的不断普及,个人隐私和安全相关协议、条款的不断完善,个人信息管理工具将会被更多的用户接受并带来全新的使用体验。

#### 注释

- [1] Vannevar Bush. As we may think. ACM, 1979 (01): 36~44
- [2] The PIM Workshop An NSF—Sponsored Invitational Workshop on Personal Information Management <http://pim.ischool.washington.edu/pim-05/home.htm>, 2009—12—21
- [3] William Jones. Keeping Found Things Found: the study and practice of personal information. America. Morgan Kaufmann publications, 2008: 4—6
- [4] Ofer Bergman, Ruth Beyth—Marom, Rafi Nachmias The Project Fragmentation Problem in Personal Information Management <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.100.6405>, 2010—4—29
- [5] Paul Alexandru Chirita, Rita Gavriloaie, Stefania Ghita Activity Based Metadata for Semantic Desktop Search <http://www.kbs.uni-hannover.de/Arbeiten/Publicationen/2004/DesktopSearch.pdf>, 2010—4—21
- [6] Jens Dietrich, Nathan Jones, Jevon Wright Using social networking and semantic web technology in software engineering—Use cases, patterns, and a case study. <http://www.computer.org/portal/web/csdl/doi/10.1109/ASWEC.2007.51>, 2010—04—28
- [7] Fausto Giunchiglia. Contextual Reasoning. Epistemologia, 1993 (16): 345—364
- [8] Patrick Brézillon. Role of Context in Social Networks. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.80.380&rep=rep1&type=pdf>, 2010—04—18
- [9] 岳玮宁,王悦,汪国平等.基于上下文感知的智能交互系统模型.计算机辅助设计与图形学学报,2005(1):74—79
- [10] J Indratmo, J Vassileva A Review of Organizational Structures of Personal Information Management Journal of digital information, 2008 (9): 1—19
- [11] Enrico Minack, Raluca Paiua, Stefania Costache et al. Leveraging personal metadata for Desktop search: The Beagle++ system. Elsevier, 2009 (12): 37—54

- 的设计. 大学图书馆学报, 2004 (1): 50-55
- [5] 高杨, 李幼平. UCL 理论及其系统设计. 电视技术, 2001 (2): 38-41
- [6] 邢玲, 史杏荣. 基于 UCL 的网页自动标引技术. 计算机工程与应用, 2004 (17): 148-151
- [7] 谢晓锋, 张文俊, 阮骏等. 针对带约束非线性规划问题的遗传算法. 计算机工程与应用, 2002 (21): 64-67
- [8] 田苗苗, 许建潮, 汪津, 丁桂英. 基于遗传算法的 Web 信息自动标引研究. 吉林大学学报 (信息科学版), 2006, (9): 542-547
- [9] 邢文训, 谢金星. 现代优化计算方法. 北京: 清华大学出版社, 1999: 140-189
- [10] WHITLEY D. The GenitorAlgorithm and Selection Pressure//Proceeding of the Third International Conference on Genetic Algorithms (ICGA-89). California: Morgan Kaufmann, 1989: 116-121
- [11] Grefenstette J J. Optimization of Control Parameters for

Genetic Algorithms IEEE Transactions on Systems, Man and Cybernetics, 1986, 16 (1): 122-128

参考文献

- 1 侯汉清, 薛鹏军. 基于知识库的网页自动标引和自动分类系统的设计. 大学图书馆学报, 2004 (1)
- 2 邢玲, 史杏荣. 基于 UCL 的网页自动标引技术. 计算机工程与应用, 2004 (17)
- 3 田苗苗, 许建潮, 汪津, 丁桂英. 基于遗传算法的 Web 信息自动标引研究. 吉林大学学报 (信息科学版), 2006 (9)
- 4 郭利伟, 王家兵. Web 信息自动标引方法比较. 内蒙古科技与经济, 2009 (7)

陈桂菊 安徽大学管理学院图书馆学专业硕士研究生。  
郭春侠 安徽大学管理学院硕士生导师, 副教授。

(上接第 54 页)

- [12] Leo Sauermann, Gunnar Grimnes, Thomas Roth-Berghofer. The Semantic Desktop as a foundation for PIM research. <http://pim2008.ethz.ch/papers/pim2008-sauermann-etal.pdf>, 2010-04-21
- [13] Yukun Li, Xiaofeng Meng. Research on Personal Dataspace Management. <http://portal.acm.org/citation.cfm?id=1410311>, 2010-04-21
- [14] Duen Horng Chau, Brad Myers, Andrew Faulring. What to Do When Search Fails: Finding Information by Association. CHI 2008: 999-1000
- [15] Khalid Latif, A Min Tjoa. Combining Context Ontology and Landmarks for Personal Information Management. <http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=05276629>, 2010-04-11
- [16] David Huynh, David Karger, Dennis Quan. Haystack: A Platform for Creating, Organizing and Visualizing Information Using RDF. <https://www.aaai.org/Papers/Workshops/2002/WS-02-11/WS02-11-014.pdf>, 2010-5-13
- [17] Sven Schwarz. A Context Model for Personal Knowledge Management. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.94.6697&rep=rep1&type=pdf>, 2010-4-21
- [18] Yuhua Cai, Xin Luna Dong, Alon Halevy, Jing Michelle Liu, and Jayant Madhavan. Personal Information Management with Semex. SIGMOD. 2005: 921-923

刘芳 中国科学院国家科学图书馆 2008 级硕士研究生。  
李春旺 中国科学院国家科学图书馆研究馆员。