

可视化技术在竞争情报中的应用

孙洁丽^{1,2,3}, 景民昌^{1,2}

(1. 中国科学院文献情报中心, 北京 100080; 2. 中国科学院研究生院, 北京 100080;
3. 河北经贸大学信息技术学院, 河北 石家庄 050000)

摘要: 在简要介绍信息可视化技术后, 分析了几种可视化技术和模型, 对所涉及的关键技术进行了较为详细的讨论。根据竞争情报的特点, 并借助于迅速发展信息可视化思想和技术, 提出了基于信息可视化技术的竞争情报系统的设计, 着重分析了可视化竞争情报系统的五个模块, 包括情报收集模块、情报文档分析模块、可视化数据处理模块、可视化显示模块和数据导入及转换模块。

关键词: 信息可视化; 竞争情报; 用户视图; 数据处理

中图分类号: G350.7; TP39

文献标识码: A

文章编号: 1007-7634(2007)07-1089-04

The Application of Visual Technology in Competitive Intelligence

SUN Jie-li^{1,2,3}, JING Min-chang^{1,2}

(1. Library of Chinese Academy of Sciences, Beijing 100080, China;

2. Graduate University of Chinese Academy of Sciences, Beijing 100049, China;

3. College of Information Technology, Hebei University of Economics and Business, Shijiazhuang 050061, China)

Abstract: This paper first introduces several models of information visualization, and discusses their key technologies particularly. Then, according to the characterises of CI, a frame of CI system based on visual technologies is put forwards, which comprises five modules: sources collecting, document analyzing, data processing, result visualizing, importing and transforming.

Key words: information visualization; competitive intelligence; user view; data processing

1 引言

竞争情报 (Competitive Intelligence, 简称 CI) 是关于竞争环境、竞争对手、竞争态势和竞争策略的信息和研究。随着信息技术的发展, 竞争情报的获取、生产和传播通过竞争情报系统 (Competitive Intelligence System, 简称 CIS) 来实现^[1-2]。

竞争情报系统除了需要对海量数据进行存储、传输、检索及分类外, 更需要迫切了解数据之间的相互关系及发展趋势。实际上, 在激增的数据背

后, 隐藏着许多重要的信息, 人们希望能够对其进行更高层次的分析, 以便更好地利用这些数据。为了了解数据之间的相互关系及发展趋势, 人们可以借助于信息可视化技术。

2 可视化技术与模型的选择

所谓信息可视化就是利用计算机支撑的、交互的、对抽象数据的可视化表示, 来增强人们对信息集合, 如 Web 数据、商业信息、文档集合、抽象概念等非物理抽象信息的认知。信息可视化技术的

收稿日期: 2007-04-25

基金项目: 河北省科技厅科研课题(042135121)

作者简介: 孙洁丽(1969-), 女, 山东省文登人, 副教授, 博士生, 从事文本挖掘、可视化与竞争情报系统研究; 景民昌(1972-), 男, 山西省永济人, 工程师, 博士生, 从事可视化技术与竞争情报系统。

核心是把文献信息、用户提问、各类情报信息检索模型以及用情报检索模型进行信息检索的过程中不可见的内部语义关系转换成图形,展示在一个低维的可视化空间中,提供一种有效的信息反馈机制^[3]。

信息可视化被理解为一种计算、一种转换,但是从信息服务角度它更应该是一个界面。无论信息系统内部是什么样的结构,对于用户而言,或者对于信息提供而言,需要的是以简洁易懂,省时高效方式表示信息内容。信息可视化技术不仅在揭示信息资源的广度与深度上有很大的优势,而且它能够将隐藏在信息资源内部的、复杂的、抽象的语义以直观的图形方式呈现给用户^[4-5]。

研究信息可视化,最终的目标是如何提供可视化的信息服务。信息可视化不仅用图像来显示多维的非空间数据,使用户加深对数据含义的理解,而且用形象直观的图像来指引情报检索过程,加快情报检索速度。因此开发可视化竞争情报系统是十分迫切和必要的。

面对不断膨胀的信息资源,可视化竞争情报系统具有广泛的应用前景。可视化竞争情报系统的关键问题之一是选择合适的可视化技术和模型,合理表示出信息资源的内在关系。

下面对几种常见的可视化技术和模型做分析。

(1) Focus + Context (焦点 + 上下文) 技术。Focus + Context 技术以忽略细节的方式显示尽可能多的周围信息,将周围信息以及以细节方式显示的焦点信息结合在一起,随着用户注意的变化改变细节显示部分。这种技术可以将一个信息集合的特定部分的细节视图,通过某种方式和该信息集合的总体结构视图混合在一起;也可以认为是在显示一个大的信息空间 (Context) 的同时,其中的一部分以更细节的方式显示 (Focus)^[6-7]。

Focus + Context 技术的核心问题是保证 Focus 信息正常显示的前提下,怎样才能显示更多的 Context 信息,以及 Focus 区域与 Context 区域的方便切换。

(2) Hyperbolic Tree 模型。Hyperbolic Tree 是一种基于双曲几何的可视化和操纵大型层次结构的 Focus + Context 技术。这种技术将更多的可视化空间用于显示层次结构中当前被关注的部分,而同时又能把整个层次结构显示出来。它通过一种规范的算法将层次关系显示在一个双曲平面上,然后将这个双曲平面映射到显示区域中。双曲线树因它的新颖、动态和实际的应用而引起了很多关注。例如:

美国国家科学数字图书馆 (NSDL) 的 “NSDL at a Glance”, 它提供了一个双曲线树,且允许用户与层次结构进行交互。用户能动态地集中注意力到这个层次结构的任何特殊部分以优化他们的观察^[8]。

Hyperbolic Tree 通过方便的交互手段很好的解决了层次结构中 Focus + Context 之间的平滑过渡问题。Hyperbolic Tree 的缺点是把当前关注的节点移到显示区域的中心时,其父节点则被推移到周围,这时将很难分辨出父子关系;难以体现同一层次上不同父亲的子节点的关系。

(3) Tree - map 模型。Tree - map 是一种表示层次信息的可视化模型。在这种结构中,层次结构的每个节点被表示为一个矩形,矩形的面积表示相应节点的权重。Tree - map 将显示空间分割成互相包围的一些矩形,这些矩形表示树形结构。包围在某个矩形中的节点的画法完全依赖于节点的内容,每个节点的显示尺寸基于它相对于整棵树的比重。表示同一父节点的所有子节点的矩形被该父节点的矩形包围着。Tree - map 可视化结构充分利用了显示空间,通过一种空间填充策略将层次结构映射为矩形。这种对空间的有效利用使大型层次结构能够在有限的空间中进行显示,并且使语义信息的表示变得相对容易了^[9]。

Tree - map 的优点是可以有效利用计算机屏幕空间,并且易于实现。它的缺点是丧失了层次结构的直观性。并且也丧失了对处于同一层次上不同父亲的子节点的关系表示,而这种关系对于把握节点之间的层次关系的结构特征是非常有用的。

(4) Cone Tree 模型。Cone Tree 采用 3D 技术,对层次结构进行可视化表示,其基本思想是利用三维图形技术将传统的二维树形表示法扩展到三维空间。在 Cone Tree 中,表示层次结构的整棵树以三维的形式进行组织和显示,每个节点和它的子节点布局呈锥形,利用三维锥形体来实现层次结构中父子节点之间的连接。为了让用户可以观察到所有的数据,锥形是半透明的,而且辅以旋转、拖动等交互技术,层次树可以转动。因此,可以实现对复杂层次关系的把握^[10]。

Cone Tree 的优点主要有:可以比较容易的体现出树形结构的整体信息;在有限的屏幕空间中可以显示更多的节点;可以利用更多的手段来提供信息,如除了几何结构以外,还可以通过圆锥在平面上的阴影映射显示节点的分布情况。

Cone Tree 的缺点主要有:空间分布不均匀,

有的区域信息过于密集，而有的空间则完全没有得到利用，尤其是在层次结构本身不平衡的情况下尤为突出；节点之间的相互遮挡情况较严重。

可视化竞争情报系统的设计和实现首先需要确定可视化对象，并根据可视化对象间的关系抽象出虚拟结构。

确定可视化对象是指分析情报数据集的内部特点（语义结构、链接关系、引用关系等），抽取出结构化信息（如链接、引用关系）、上下文信息、元数据（文件大小、作者）、使用信息（如浏览顺序、操作）及语义信息（领域、关键词、摘要、标题等）等，为选择合适的可视化技术和模型提供依据。

其次是将抽象信息特征映射成空间化、图形化的特征，即完成信息的可视化映射；然后根据映射的结果选用合适的可视化的技术和模型对可视化对象进行组织，构建可视化空间；最后完成可视化图形的绘制，将最终的可视化结果呈现给用户。

3 可视化竞争情报系统的设计

可视化竞争情报系统在软件功能上应该给用户提供最大的方便，能够使用户以最直观的方式获得需求信息。随着信息规模的急剧增长，这一点显得尤为重要，在激增信息数据信息背后，隐藏着许多重要信息，用户希望能够对其进行更高层次的分析，以便更好地利用数据。因此把信息可视化技术引入到竞争情报系统中受到用户的关注。

3.1 可视化竞争情报系统用户视图

可视化竞争情报系统功能的主要目的是为用户提供更直观的方式显示情报数据，并利用可视化功能将情报数据之间的关系转换成图形，在一个二维或三维的可视化空间中显示出来。

可视化竞争情报系统的用户视图如图1所示。

3.2 可视化竞争情报系统的总体结构

信息可视化被理解为一种计算、一种转换，但是从信息服务角度它更应该是一个界面。无论信息系统内部是什么样的结构，对于用户而言，或者对于信息提供而言，需要的是以简洁易懂，省时高效方式表示信息内容。因此研究信息可视化，最终的目标是如何提供可视化的信息服务。

实现可视化情报信息提取的主要过程为：首先规划和定向，即确立竞争情报过程的目标；其次是

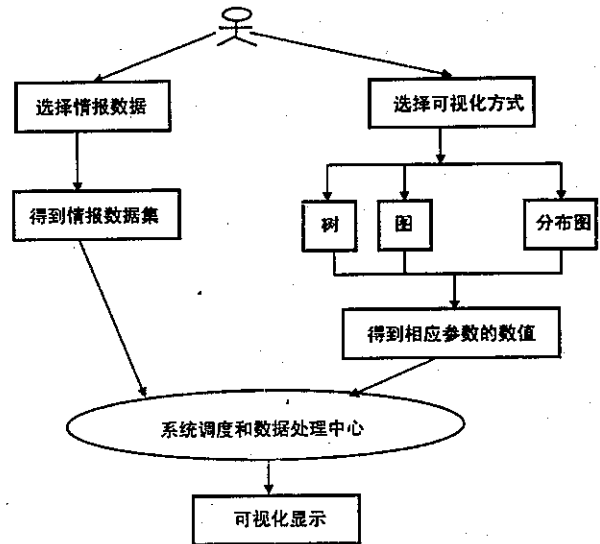


图1 可视化竞争情报系统用户视图

信息收集，即采用合适的信息收集方法，收集企业内外的相关信息；再次是信息加工，即初步整理收集来的数据和信息，使之有序化；然后是情报分析，对加工后的信息进行深入分析，从而产生出竞争情报；最后是情报服务。将情报产品以最有效的方式展现给情报用户，为其决策提供依据，并接受用户的反馈信息，以提高竞争情报工作的质量。

根据以上过程，可以给出可视化竞争情报系统的逻辑结构图，如图2所示。

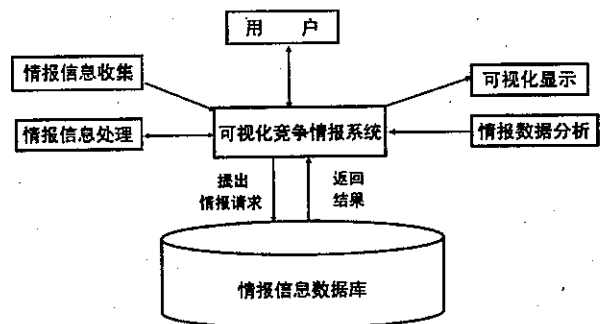


图2 可视化竞争情报系统的逻辑结构图

3.3 可视化竞争情报系统的基本功能和主要模块

可视化竞争情报系统可以划分为五个模块：情报收集模块、情报文档分析模块、可视化数据处理模块、可视化显示模块和数据导入及转换模块。

(1) 情报收集模块。情报收集模块主要完成的功能是根据用户的设定，借助元搜索引擎，进行情报信息收集，得到与用户设定主题相关的情报信

息,并对不同搜索引擎的搜索结果进行融合,得到一个包含多种格式的情报文档集,为下一步的情报文档预处理提供原始的信息。该模块不是完全由自我实现的搜索引擎,而是采用针对竞争情报主题的专题化搜索引擎,并结合元搜索引擎技术,使竞争情报关键字的检索结果达到更高的相关度。

(2) 情报文档分析模块。将前面得到的与搜索主题相关的文档集中不同的文档格式转换成统一的格式文件,然后通过文档分析技术对统一的格式文件进行分析处理,以获得文档摘要和文档结构信息,对得到的文档集和文档结构信息进行竞争情报的分析,如可以利用相关的数据挖掘和文本挖掘技术进行基于关键字的关联挖掘、文档的分类和聚类及其分析,得到竞争情报信息集合。

(3) 可视化数据处理模块。可视化数据处理模块的主要功能是为实现可视化图形模块生成基础数据。该模块对相关竞争情报信息集合进行处理。通过可视化数据处理模块可以为可视化显示准备好数据。

(4) 可视化显示模块。可视化显示模块是系统的核心模块,该模块的主要功能是实现可视化显示,使用可视化显示模块可以浏览相应情报信息的直观状态。本模块使用相应的可视化算法将加入图形属性数据后的竞争情报信息用合适的可视化图形展现出来,并按用户要求对图形做进一步操作或者保存所得图形。

(5) 数据导入及转换模块。本模块主要用于系统基础数据的生成。数据导入功能是从现有的竞争情报信息数据库中,提取数据集,作为该系统的基础数据。数据转换功能用于转换数据。由于不同的竞争情报信息数据库,信息数据的格式、信息数据的字段定义一般都不相同,把不同格式的数据以半自动转换和自动转换相结合的方法,转换为统一的数据格式,使竞争情报系统独立于信息资源数据库,具有灵活的系统接口。

4 结 语

制约我国竞争情报服务发展的重要因素之一是信息服务创新不够。具备良好的服务创新能力才能提高服务内容的质量,从而准确地把握用户的需求。可视化的竞争情报系统运用现代信息技术和可视化技术,对情报信息进行存贮、加工和处理,以图形交互界面为主要特色,与传统的竞争情报系统

相比,可视化的竞争情报系统能够采用更加直观形象的人机交互并对情报数据进行多种加工处理,从而在情报决策方法、方式、效率和质量等方面前进了一大步^[11-12]。

参考文献

- 1 杨少萱. 国外竞争情报的发展与应用[J]. 中国信息导报, 2006, (11): 28-30.
- 2 吴晓伟, 宋文官, 徐福缘. 竞争情报软件发展现状和趋势研究[J]. 情报杂志, 2006, 25(6): 2-5.
- 3 张学福. 基于知识模型的文本信息检索可视化研究[J]. 中国图书馆学报, 2006, 32(5): 52-56.
- 4 周宁, 张玉峰, 张李义. 信息可视化与知识检索[M]. 北京: 科学出版社, 2005: 20-28.
- 5 董献洲, 胡晓峰, 司光亚. 信息可视化技术在情报分析中的应用研究[J]. 计算机工程与应用, 2006, 42(34): 175-177.
- 6 San Diego, California. TreeJuxtaposer: scalable tree comparison using Focus + Context with guaranteed visibility[J]. ACM Transactions on Graphics (TOG), 2003, 22(3): 453-462.
- 7 Matthias Kreuzler, Heidrun Schumann. Information visualization using a new focus + context technique in combination with dynamic clustering of information space[Z]. Proceedings of the 1999 workshop on new paradigms in information visualization and manipulation in conjunction with the eighth ACM international conference on Information and knowledge management: 1-5. 2007-01-25.
- 8 Peter Piroli, Stuart K. Card, Mija M. Van Der Wege. The effects of information scent on visual search in the hyperbolic tree browser[J]. ACM Transactions on Computer-Human Interaction (TOCHI), 2003, 10(1): 20-53.
- 9 Ben Shneiderman. Tree visualization with tree-maps: 2-d space-filling approach[J]. ACM Transactions on Graphics (TOG), 1992, 11(1): 92-99.
- 10 George G. Robertson, Jock D. Mackinlay, Stuart K. Card. Cone Trees: animated 3D visualizations of hierarchical information[Z]. Proceedings of the SIGCHI conference on Human factors in computing systems: Reaching through technology: 189-194, 1991. 2007-01-25.
- 11 Hwee-Leng Ong, Ah-Hwee Tan, Jamie Ng, Hong Pan, Qiu-Xiang Li. Knowledge Management: Organizing What You Know: FOCI: flexible organizer for competitive intelligence[Z]. Proceedings of the tenth international conference on Information and knowledge management: 523-525, October 2001. 2007-01-25.
- 12 余玄霞. 我国竞争情报服务技术创新分析[J]. 情报科学, 2006, 24(12): 1878-1882.

(责任编辑: 刘莹)